DOCTOR OF PHILOSOPHY

Genomic investigation into the evolution of Staphylococcus aureus and the micro-epidemiology of Staphylococcus aureus in atopic eczema

Harkins, Catriona P.

Award date:
2016

Genomic investigation into the evolution of *Staphylococcus aureus* and the micro-epidemiology of *Staphylococcus aureus* in atopic eczema

Catriona P. Harkins

School of Life Sciences

University of Dundee

2016

Table of Contents

List of Figures ..... **vi**
List of Tables ..... **viii**
Abbreviations ..... **xi**
Glossary ..... **xii**
Acknowledgements ..... **xiv**
Declaration ..... **xv**
Summary ..... **xvi**

1 Introduction **1**
1.1 *Staphylococcus aureus* **2**
1.1.1 Diversity of the *S. aureus* population 3
1.1.2 The *S. aureus* genome 4
1.1.3 Niche adaptation 10
1.2 Variation occurring in the genome over short time scales **12**
1.2.1 Within host diversity during disease 12
1.2.2 Between host variability 14
1.2.3 Bacterial Adaptation to the host 14
1.3 The evolution of antimicrobial resistance in *S. aureus* **25**
1.3.1 Timeline of emerging antimicrobial resistance in *S. aureus* 25
1.3.2 Beta-lactamase resistance in *S. aureus* 29
1.3.3 The emergence of the first MRSA and beyond 32
1.4 Atopic eczema and its association with *S. aureus* **35**
1.4.1 Atopic eczema: Prevalence and pathogenesis 35
1.4.2 Enhanced colonisation by *S. aureus* in AE 39
1.4.3 How does *S. aureus* contribute to disease activity in AE? 41
1.4.4 Bacterial factors and clinical epidemiology 44
1.5 Other microbes in AE **48**
1.6 Overall aims of the study **49**

2 Materials and methods **51**
2.1 Study design **52**
2.1.1 Ethical and local authority approval 52
2.1.2 Recruitment criteria 53
2.2 Media and reagents **53**
2.2.1 Media and bacterial growth conditions 53
2.2.2 Reagents 54
2.3 Isolation and identification of *S. aureus* from clinical samples **55**
2.3.1 Swabs and sampling 55
2.3.2 Isolation of *S. aureus* from clinical swabs 55
2.3.3 Colony PCR 56
2.4 DNA methods **57**
2.4.1 Extraction 57
2.4.2 Quantification and quality assessment of DNA 58
2.5 Whole genome sequencing (WGS) **58**
2.5.1 DNA library preparation 58
2.5.2 Sequencing 60
2.6 Bioinformatic analysis of sequencing data **60**
2.6.1 Assembly and mapping 61
2.6.2 Multi-locus sequence typing 65
2.6.3 Phylogenetic reconstruction 65
2.6.4 SNP reconstruction 70
2.6.5 Insertion/deletion analysis 70

2.6.6 Genome annotation 71
2.6.7 Recombination analysis 71
2.6.8 Temporal and mutation rate analysis 71
2.6.9 Virulence determinant profiling 72
2.6.10 Functional gene classification 77
2.6.11 Pangenome analysis 78
2.6.12 Discriminant analysis of the principle components (DAPC) 78
2.7 Statisical analysis 78

3 Genomic investigation of the emergence of the first Methicillin resistant *Staphylococcus aureus* (MRSA) 80
3.1 Introduction 81
3.2 Aims and Objectives 82
3.3 Origins of the Historic MRSA collection 82
3.4 Population structure of the historic collection 83
3.5 Genetic basis of antimicrobial resistance within the population 88
3.5.1 Resistance genotype of the population 88
3.5.2 The SCC*mec* element in the historic collection 91
3.6 Investigating the evolutionary events leading to emergence of the first MRSA 92
3.6.1 The acquisition of SCC*mec* 92
3.6.2 Estimating the age of the Archaic MRSA clone 95
3.6.3 Single acquisition of an SCC*mec* element leading to the emergence of MRSA 99
3.7 Discussion 102
3.8 Concluding remarks 107

4 *Staphylococcus aureus* carriage in healthy children 109
4.1 Introduction 110
4.2 Aims and objectives 110
4.3 Recruitment, participation and colonisation rates 111
4.3.1 Sampling method and rationale 113
4.4 Clinical phenotype of recruits 113
4.5 Isolation of *S. aureus* from clinical samples 115
4.5.1 Microbiological burden 115
4.5.2 Confirmation of *S. aureus* from clinical samples 116
4.6 Genetic background of colonising strains 116
4.6.1 ST30 comparison 117
4.7 Resistome profile of the colonising populations 119
4.8 Genetic heterogeneity in *S. aureus* colonising healthy children 120
4.8.1 Diversity of colonisation 120
4.8.2 Estimating the age of the colonising population 121
4.8.3 Diversity within individuals 123
4.8.4 Potential relevance of diversity 131
4.9 Discussion 133
4.10 Concluding remarks 138

5 *Staphylococcus aureus* associated with Atopic Eczema prospective case study .. 139
5.1 Introduction 140
5.2 Aims and objectives 140
5.2.1 Recruitment 141
5.2.2 Sampling method and rationale 141
5.3 Clinical phenotype of recruits 142
5.4 AE disease severity and *S. aureus* burden from clinical samples 143
5.5 Colonising strains 144
5.6 Resistome profile of AE associated strains 145
5.7 Genetic heterogeneity within individuals 147

5.7.1 Diversity of colonising *S. aureus* 147
5.7.2 Estimating the age of the colonising population of *S. aureus* in AE cases 148
5.7.3 Diversity within individual AE cases 150
5.8 Comparison of colonisation dynamics and heterogeneity between cases and controls 172
5.8.1 Diversity and duration of colonisation in cases and controls 172
5.8.2 Differentiation of *S. aureus* during colonisation: case control comparison 175
5.9 Discussion 179
5.9.1 Diversity of *S. aureus* in atopic eczema disease flares 179
5.9.2 Diversity in *S. aureus* colonisation between nasal carriage controls and AE cases 183
5.10 Concluding remarks 186
5.10.1 Diversity of S. aureus colonising children with AE 186
5.10.2 Diversity of *S. aureus* carriage in healthy children versus AE cases 187

6 *Staphylococcus aureus* associated with Atopic Eczema: case control study 188
6.1 Introduction 189
6.2 Aims and objectives 189
6.3 Recruitment criteria and sampling 190
6.4 Clinical phenotype of recruits 191
6.5 Clonal complex breakdown between cases and controls 192
6.6 Comparison of virulence gene content between disease and carriage isolates 194
6.6.1 Toxins and immune evasion determinants 195
6.6.2 Antimicrobial resistance genes 201
6.6.3 Adhesion genes 206
6.7 Pangenome comparison of disease and carriage isolates 211
6.8 Discussion 218
6.9 Concluding remarks 225

7 Conclusions and outlook 227
7.1 The selective pressures of clinical practice and the emergence of MRSA 228
7.2 *S. aureus* diversity and adaptation during colonisation 229
7.3 Differences in *S. aureus* populations carried by AE cases and nasal controls 232

8 References 234

Appendix A: Chapter 4 Supplementary data 253
Control study participant SS_016 254
Control study participant SS_026 255
Control study participant SS_045 256
Control study participant SS_057 257
Control study participant SS_059 258
Control study participant SS_091 259
Control study participant SS_094 260
Control study participant SS_099 261
Control study participant SS_141 262
Control study participant SS_147 263
Control study participant SS_153 264
Control study participant SS_157 265
Control study participant SS_250 266
Control study participant SS_268 267
Control study participant SS_303 269

Appendix B: Chapter 5 Supplementary data 270
Patient 2 (study ID PSAE002) 271

Patient 3 (study ID PSAE003) 272
Patient 6 (study ID PSAE006) 274
Patient 9 (study ID PSAE009) 276
Patient 10 (study ID PSAE010) 278

List of Figures

Figure 1-1 In host diversity and adaptation in *S. aureus* 21
Figure 1-2 Antimicrobial introduction from 1940 onwards and the emergence of resistance in *S. aureus* 27
Figure 1-3 *SCCmec* components 32
Figure 1-4 The human skin barrier 37
Figure 3-1 Population structure of the historic collection 85
Figure 3-2 Jevons Isolates within the Historic collection 87
Figure 3-3 Resistance genotype of the historic collection 90
Figure 3-4 Recombination within SCC*mec* type I elements of the historic collection 93
Figure 3-5 Maximum likelihood core SNPs tree SCC*mec* type I elements in historic collection 94
Figure 3-6 Root to tip distance regression analysis using TempEst 96
Figure 3-7 SCC*mec* type I haplotypes in the historic collection and their distribution within the population 101
Figure 4-1 Community control study recruitment flow chart 112
Figure 4-2 *femB* Colony PCR confirming *S. aureus* from clinical samples 116
Figure 4-3 ST30 isolate comparison across recruitment sites 118
Figure 4-4 In-host diversity levels in nasal carriage controls 121
Figure 4-5 Estimated age of the colonising population in asymptomatic carriage controls 122
Figure 4-6 Maximum likelihood core SNPs tree for control study ID SS_039 124
Figure 4-7 Maximum likelihood core SNPs tree for control study ID SS_105 126
Figure 4-8 Maximum likelihood core SNPs tree for control study ID SS_149 129
Figure 5-1 In-host diversity levels across AE cases 148
Figure 5-2 Estimated age of the colonising population in Atopic eczema cases 149
Figure 5-3 Maximum likelihood core SNPs for Patient 1 (study ID PSAE001) 153
Figure 5-4 Evidence of adaptive mutations in Patient 4 (study ID PSAE004) 161
Figure 5-5 Evidence of section persistence within the host in patient 5 (study ID PSAE005) 167
Figure 5-6 Colonisation by two distinct STs in Patient 8 (study ID PSAE008) 171
Figure 5-7 Diversity and colonisation duration in cases and controls 174
Figure 5-8 Comparison of SNP types between cases and controls 176
Figure 5-9 Distribution of non-synonymous base substitutions across the *S. aureus* genome by functional gene classes 178
Figure 6-1 Clonal complex breakdown of isolates from age-matched AE cases and nasal carriage controls 194
Figure 6-2 Immuno-modulatory gene prevalence in cases and controls 198
Figure 6-3 Immuno-modulatory gene prevalence in cases and controls according to clonal complex 200
Figure 6-4 Prevalence of antimicrobial resistance determinants across case and control populations 204
Figure 6-5 Mapping against reference allele of *eap* (extracellular adherence protein) from CC8 reference genome NCTC8325 (locus ID: SA0UHSC_02161) in Irish AE cases 208
Figure 6-6 Assessment of genome assemblies and number of genes in pangenome 214
Figure 6-7 Discriminant analysis of Principal components applied to 93

S. aureus isolates 215
Figure 6-8 Accessory genes differentiating case and control population 216

Supplementary Figures:

Figure S1 Maximum likelihood core SNPs tree for control study ID SS_016 254
Figure S-2 Maximum likelihood core SNPs tree for control study ID SS_026 255
Figure S3 Maximum likelihood core SNPs tree for control study ID SS_045 256
Figure S-4 Maximum likelihood core SNPs tree for control study ID SS_057 257
Figure S-5 Maximum likelihood core SNPs tree for control study ID SS_059 258
Figure S-6 Maximum likelihood core SNPs tree for control study ID SS_091 259
Figure S-7 Maximum likelihood core SNPs tree for control study ID SS_094 260
Figure S-8 Maximum likelihood core SNPs tree for control study ID SS_099 261
Figure S-9 Maximum likelihood core SNPs tree for control study ID SS_141 262
Figure S-10 Maximum likelihood core SNPs tree for control study ID SS_147 263
Figure S-11 Maximum likelihood core SNPs tree for control study ID SS_153 264
Figure S-12 Maximum likelihood core SNPs tree for control study ID SS_157 265
Figure S-13 Maximum likelihood core SNPs tree for control study ID SS_250 266
Figure S-14 Maximum likelihood core SNPs tree for control study ID SS_268 267
Figure S-15 Maximum likelihood core SNPs tree for control study ID SS_303 269
Figure S-16 Maximum likelihood core SNPs tree for Patient 2 (PSAE002) 271
Figure S-17 Maximum likelihood core SNPs tree for Patient 3 (PSAE003) 272
Figure S-18 Maximum likelihood core SNPs tree for Patient 6 (PSAE006) 274
Figure S-19 Maximum likelihood core SNPs tree for Patient 9 (PSAE009) 276
Figure S-20 Maximum likelihood core SNPs tree for Patient 10 (PSAE010) 278

List of Tables

Table 2-1 Reagents used for this study....54
Table 2-2 Primers used in this study....57
Table 2-3 *S. aureus* reference genomes used for in this study. 63
Table 2-4 *S. aureus* genome assemblies used for self-mapping....64
Table 2-5 Accessory regions masked from MSSA476 reference chromosome 66
Table 2-6 Accessory regions masked for N315 reference chromosome. 66
Table 2-7 Accessory regions masked from COL reference chromosome 67
Table 2-8 Accessory regions masked from HO 5096 0412 (EMRSA15) reference chromosome....67
Table 2-9 Accessory regions masked from MRSA252 reference chromosome. 68
Table 2-10 Accessory regions masked from MO13 (ST59) reference chromosome....69
Table 2-11 Accessory regions masked from CUHK_HK188 (ST188) reference chromosome. 69
Table 2-12 Accessory regions masked from CA347 (ST45) reference chromosome. 70
Table 2-13 *S. aureus* toxin reference alleles used for this study. 74
Table 2-14 Clonal complex specific *S. aureus* adhesion gene alleles used for this study. 75
Table 2-15 Antimicrobial resistance determinant alleles used in this study. 76
Table 2-16 Point mutations conferring antimicrobial resistance assessed for in this study. 77
Table 3-1 Predicted regions of recombination from Gubbins analysis. 84
Table 3-2 Minimum inhibitory concentration (MIC) to Methicillin (Celbenin) derived from the original Jevons description, BMJ 1961. 88
Table 3-3 Homoplastic SNPs identified in core genome alignment of 122 isolates used for temporal analysis....97
Table 3-4 Canonical SNPs in sub-group of 120 SCC*mec* type I elements, derived between 1960 and 1968. 100
Table 4-1 Nasal colonisation control characteristics 114
Table 4-2 Bacterial burden in nasal carriage controls 115
Table 4-3 Colonising strains in nasal carriage controls....117
Table 4-4 Resistance profiles of community carriage control isolates 120
Table 4-5 Size of core genome from representative reference genomes and self- assemblies used for mapping and SNP assessment....122
Table 4-6-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_039 125
Table 4-6-2 Within host-heterogeneity identified in nasalcarriage control Study ID SS_039 125
Table 4-7-1 Within host-heterogeneity identified in nasal carriage of control individual ID SS_105 127
Table 4-7-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_105....128
Table 4-8-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_149....130
Table 4-8-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_149. 130
Table 5-1 Atopic eczema case characteristics. 143
Table 5-2 Correlation of disease severity and number of body sites colonised by *S. aureus* in AE cases. 143
Table 5-3 Bacterial burden in AE cases. CFU counts from 9 cases sampled across 5 body sites....144
Table 5-4 Colonising strains in AE cases. Sequence type (ST), clonal complex (CC), and MLST profile is representative of all sequenced colonies for the individual 145

Table 5-5 Resistance profiles of AE case isolates. 147
Table 5-6 Size of core genome from representative reference genomes and self assemblies used for mapping and SNP assessment. 149
Table 5-7 Within host-heterogeneity identified in Patient 1(study ID: PSAE001). 154
Table 5-8 Within host-heterogeneity identified in Patient 1 (study ID: PSAE001). 159
Table 5-9 Within host heterogeneity identified in Patient 4 (study ID: PSAE004). 162
Table 5-10 Within host-heterogeneity identified in Patient 4 (study ID: PSAE004). 165
Table 5-11 Within host heterogeneity identified in Patient 5 (study ID: PSAE005). 168
Table 5-12 Within host-heterogeneity identified in Patient 5 (study ID: PSAE005). 169
Table 5-13 Within host-heterogeneity identified in Patient 8 – ST123 colonies (study ID: PSAE008). 172
Table 6-1 Atopic eczema case and control characteristics. 192
Table 6-2 Chi2 comparison of colonising strain backgrounds in cases and controls 193
Table 6-3 Chi2 comparison of immune-modulatory gene prevalence in cases and controls. 197
Table 6-4 Chi2 comparison of antimicrobial resistance determinant prevalence between cases and controls 203
Table 6-5 Chi2 comparison of Fusidic acid resistance determinant prevalence between cases and controls 205
Table 6-6 Chi2 comparison of antiseptic resistance determinant prevalence between cases and controls 206
Table 6-7 S. aureus surface protein profile of Irish AE cases and controls 210
Table 6-8 Genes differentiating case and controls from DAPC analysis 217

Supplementary tables:

Table S-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_016 254
Table S-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_026 255
Table S-3-1 Within host diversity identified in nasal carriage control Study ID SS_045 256
Table S-3-2 Within host diversity identified in nasal carriage control Study ID SS_045 257
Table S-4 Within host-heterogeneity identified in nasal carriage control Study ID SS_091 259
Table S-5-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_141 262
Table S-5-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_141 263
Table S-6 Within host-heterogeneity identified in nasal carriage control Study ID SS_147 263
Table S-7 Within host-heterogeneity identified in nasal carriage control Study ID SS_153 264
Table S-8 Within host-heterogeneity identified in nasal carriage control Study ID SS_157 265
Table S-9-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_250 266
Table S-9-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_250 267
Table S-10-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_268 268
Table S-10-2 Within host-heterogeneity identified in nasal carriage control

Study ID SS_268 268
Table S-11 Within-host heterogeneity identified in Patient 2 (PSAE002). 271
Table S-12-1 Within-host heterogeneity identified in Patient 3 (PSAE003) 273
Table S-12-2 Within host-heterogeneity identified in Patient 3 (PSAE003) 273
Table S-13 Within-host heterogeneity identified in Patient 6 (PSAE006). 275
Table S-14-1 Within-host heterogeneity identified in Patient 9 (PSAE009) 277
Table S-14-2 Within-host heterogeneity identified in Patient 9 (PSAE009) 277
Table S-15-1 Within-host heterogeneity identified in Patient 10 (PSAE010) 279
Table S-15-2 Within-host heterogeneity identified in Patient 10 (PSAE010) 281

Supplementary data tables pertaining to sequencing and assembly quality for each of the studies presented in results chapters 3 to 6 are available online via:

https://figshare.com/

Ethics committee and local authority approval letters for studies described in Chapters 4 to 6 are also available online via:

https://figshare.com/

Login details

username: charkins@dundee.ac.uk
password: cpharkins2016

Abbreviations

ACT	Artemis comparison tool
AE	Atopic eczema (interchangeable with AD)
AMP	Anti-microbial peptide
AMR	Anti-microbial resistance
BHI	Brain heart infusion
bp	Base pair
BWA	Burrows-Wheeler alignment
CA	Community-acquired
CC	Clonal complex
CoN	Coagulase negative
CWA	Cell wall anchored
DAPC	Discriminant analysis of the principal components
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EASI	Eczema Area Severity Index
gDNA	Genomic DNA
GI	Genomic Island
HA	hospital acquired
IgE	Immunoglobulin E
IL	Interleukin
Indel	Insertion/ deletion
kDA	Kilo Dalton
LA	Livestock associated
LPxTG	Leucine-Proline-any-Threonine-Glycine motif
MGE	Mobile genetic element
MIC	Minimal inhibitory concentration
ML	Maximum likelihood
MLST	Multi-locus sequence type
MRSA	Methicillin resistant *Staphylococcus aureus*
MSSA	Methicillin sensitive *Staphylococcus aureus*
PBP	Penicillin binding protein
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PVL	Panton-Valentine Leuckocidin
rpm	Revolutions per minute
RNase	Ribonuclease
SaPI	*Staphylococcus aureus* pathogenicity island
SCC*mec*	Staphylococcal chromosome cassette *mec*
SE	Staphylococcal enterotoxin
Ssl	Superantigen-like
SNP	Single nucleotide polymorphism
ST	Sequence type
Tn	Transposon
Tris-EDTA	Tris- Ethylenediaminetetracetic acid
v/v	Volume per volume
w/v	Weight per volume
WGS	Whole genome sequencing

Glossary

BAM file: binary form of a SAM file which contains sequence alignment data. This is derived from mapping of sequence reads against as reference genome.

Commensal: in context of skin flora for example, living within or on a host without causing any harm.

Convergent evolution: this is the occurrence of mutations giving rise to the same phenotype in two or more unrelated, and independently evolving lineages. This can occur within the same gene, and potentially at the same site in the gene.

Facual: refers to the fauces- the oropharyngeal isthmus. Also known as the throat.

Homologous recombination: where a segment of a recipient cells genome is replaced with a homologous piece of DNA derived from the genome of the donor cell.

Homoplasy: this is a biological trait that has been acquired by species from unrelated lineages. This can refer to a structure, a phenotypic characteristic or a genetic characteristic such as a mutation. It may arise through evolutionary convergence, where the unrelated species have shared similar environments, which they have responded to by developing the same adaptive mechanism.

Horizontal gene transfer: transfer of genetic material between a recipient cell and donor cell.

Hypermutator: where an individual cell or also potentially a lineage has an increased mutational rate. This is most frequently due to loss of function of DNA repair genes.

Index primers: during WGS library preparation adapter sequences are ligated to the DNA fragments. Contained within the adapter are index primer sequences. This allows for sequencing of 24 samples (also known as multi-plexing) to be sequenced at once, given that they have a unique combination of index primer sequences. This combination can be used as a co-ordinate to then extract sequences from each individual sample, to the proceed to genome assembly.

IS element: Insertion sequence element; transposable element of short DNA sequence, usually less than 2.5 kb.

k-mer: word or small motif of length k, contained in sequence data of a genome assembly.

Mobile genetic element: segments of DNA, which can move within the genome, or between bacterial cells. Examples include plasmids and transposons.

Nosocomial: originating or being acquired from a hospital environment.

Phase variation: a bacterial mechanism to make rapid changes in a specific trait favourable to survival in response to environmental conditions. Occurs through reversible control of gene expression.

Positive selection: this is the propensity for an allele conferring either a survival or fitness advantage to the recipient to increase in frequency within a population.

Pre-adaptation: underlying yet unrecognised adaptive change that becomes evident following change in environmental conditions.

Selective pressure: this refers an environmental factor or influence that changes the behaviour and fitness of living organisms such as bacteria. These pressures drive adaptive changes, and can lead to favourable adaptive changes becoming more prevalent within a population.

Selective sweep: increasing frequency of an advantageous allele in a population, leading to its fixation in the population. This occurs as a direct result of positive selection.

Virulence: disease-causing capacity.

Acknowledgements

The work presented in this thesis has only been possible because of the collective efforts of many people. I would like to begin by thanking my supervisors. Professor Matthew Holden, who was extremely kind in allowing me to join his group at the University of St Andrews. Thank you for the huge amount of time and effort you invested in teaching me over these last two years to accomplish these projects. It has truly been a privilege to work with you, and I will always be very grateful for everything you have done to help me. To Professor Charlotte Proby who has been unwavering in her support during the course of this PhD and without whom I would have been lost. I will always look back fondly our days spent recruiting small children for nose swabs!

To the Infection Group at the School of Medicine, University of St Andrews who kindly hosted me for the last two years. It has been a pleasure working with scientists and clinicians alike that are unified in their purpose of understanding micro-organisms for the benefit of clinical practice. A very special thank you to Dr Kerry Pettigrew who spent long hours preparing DNA libraries for sequencing, forming the basis of my work, as well as always knowing when kind words and coffee were needed urgently. For her experimental input and tuition in the earliest stages of my project at St Andrews I thank Dr Katarina Oravcová and Professor Stephen Gillespie for his kind feedback and encouragement.

To my colleagues at the Dermatology Department in Ninewells Hospital, my sincerest gratitude. Professor Sara Brown for at such short notice allowing me to recruit through her Paediatric eczema clinic, as well as her careful and considered advice. To Dr Ross Hearn and Mrs Debbie Rice for their much needed assistance in recruiting to my eczema study in Ninewells. To Mrs June Gardner, who taught me how to write an ethics application on the run, and became my friend in the process. I would probably still be writing IRAS forms without you. Dr Colin Fleming and all of the consultants of the dermatology department, who have supported me to complete this work.

When desperately needed, Professor Alan Irvine and Dr Maeave McAleer kindly gifted me a sample collection that had taken them years to accumulate. That will never be forgotten. I also have to thank Dr Desiree Bennett, Professor Joan Geoghegan, Professor Tim Foster and Orla Fleury and Professor Irwin McLean for assisting the collaborative project that came of this collection. Thank you to Professor Julian Parkhill for generously allowing me to utilise the analysis tools of the Pathogen Informatics Group at the Wellcome Trust Sanger Institute to undertake all of this work.

My deepest gratitude to the children and parents who took part in these studies, as well as the head teachers and local authorities that enthusiastically participated.

Finally to my family who have without fail supported me, especially through the last three years. Your life-long encouragement to work hard and pursue knowledge is the reason I have gotten where I am today. Thank you.

Declaration

I declare that I am the author of this thesis; that unless otherwise stated, all references cited have been consulted by myself; that the work of which this thesis is a record has been done by myself and has not previously been accepted for a higher degree. Where the thesis is based upon joint research, the nature and extent of my individual contribution is defined.

Catriona P. Harkins

November 2016

Summary

Staphylococcus aureus is a highly adaptable organism, which has established itself as one the most important pathogens worldwide. It has co-evolved with humans adapting to life within the host as well as the interventions targeted to eradicate it. Despite its notoriety as a pathogen one of its natural habitats is as a commensal on the human skin. The overarching aim of this project was to understand the genetic basis of this organism's adaptation in the face of human intervention and influencing its survival in the human host, specifically during colonisation of the skin.

From humankind's perspective arguably this organism's most significant adaptation is the development of drug resistance. Methicillin resistant *Staphylococcus aureus* (MRSA) was first observed in 1960, less than one year after the introduction of the drug into clinical practice. Previous epidemiological and genetic evidence has always suggested that MRSA arose around this period, when the *mecA* gene encoding methicillin resistance carried on a Staphylococcal cassette chromosome *mec* (SCC*mec*) element, was acquired by *S. aureus*. In this work whole genome sequencing of a collection of the very first ever identified MRSA isolates was used to reconstruct the evolutionary events leading to the emergence of the archetypal MRSA lineage. This analysis revealed that *S. aureus* acquired the type I SCC*mec* element almost fourteen years prior to the first clinical use of methicillin, as a single horizontal event, and with its subsequent propagation leading to the genesis of MRSA.

Staphylococcus aureus colonisation is a characteristic feature of the inflammatory skin disease atopic eczema (AE). In AE disease exacerbations are associated with an increased burden of this pathogen. Despite this the colonisation dynamics of *S. aureus* during disease flare eczema are poorly understood. Therefore the remainder of this body of work sought to genetically interrogate AE disease-associated *S. aureus* isolates to gain further understanding of how this organism contributes to the disease pathogenesis. Firstly by characterising genetic heterogeneity arising during colonisation in periods of disease flare in children with AE, in direct comparison to healthy children asymptomatically nasally colonised, looking for evidence of micro-evolutionary change suggestive of adaptation to the host. Secondly by comparing isolates from a cohort of AE cases to childhood nasal carriers to characterise the population structure and genetic content of AE disease vs. carriage isolates. These works demonstrated that colonisation in AE during disease flare is the result of a clonal expansion of a single strain type, mirroring the colonisation dynamics found in nasal carriers. In AE cases evidence of clinically relevant adaptive mutations were identified which were linked to prolonged periods of carriage, and included examples affecting global regulators of virulence and antimicrobial resistance. This analysis also revealed segregation in the genetic backgrounds of strains preferentially colonising AE skin in comparison to nasal epithelium, and evidence of the impact of prescribing practises between the disease populations.

1 Introduction

1.1 Staphylococcus aureus

The gram-positive bacterium *Staphylococcus aureus* has established itself as one of the most important pathogens worldwide. This organism has co-evolved with humans, adapting to life within the host as well as to interventions targeted at its eradication (van Belkum et al. 2009a; Fitzgerald 2014). Its versatility as a pathogen is evidenced by the wide spectrum of illness it can cause, both in healthy and immuno-compromised hosts alike. This derives from its vast array of virulence determinants and mechanisms to manipulate the hosts immune system (Thammavongsa et al. 2015). Its potential disease manifestations vary widely from minor skin infections such as such as impetigo or cellulitis, to potentially life-threatening infections like necrotising fasciitis, pneumonia and bacteraemia (Dastgheyb and Otto 2015). Despite its notoriety as dangerous pathogen, one of its niche environments is on human skin as a commensal. Up to 60% of the population are at least intermittent nasopharyngeal carriers (Kluytmans et al. 1997; Wertheim et al. 2005), meaning its asymptomatic carriage significantly exceeds the instances of invasive disease such as bacteraemia (Public Health England 2015).

Fundamental to the success of *S. aureus* as a pathogen is its ability to adapt. Its capacity to become resistant to antimicrobials is from a global health perspective now a major concern,. At a population level its has become highly successful through its ability to adapt to differing environments such as hospitals, and wide ranging hosts from humans and across animal species (Peton et al. 2014; Fitzgerald and Holden 2016). The study of micro-evolutionary changes during its time in host, from asymptomatic carriage to invasive disease, have begun to reveal important insights on how this organism adapts to survive within individuals (Didelot et al. 2016).

1.1.1 Diversity of the *S. aureus* population

Molecular typing methods such as multi-locus sequenced typing (MLST) paved the way for our understanding of the population structure of *S. aureus*. MLST profiling involves sequencing of approximately 450bp internal fragments of 7 house-keeping genes conserved across the species (Maiden et al. 1998). Isolates are then assigned to alleles according to their sequence at these sites to generate an allelic profile. This is then used to assign the isolates to a sequence type (ST), with STs sharing 5 or more alleles being grouped into a clonal complex (CC). The genes upon which this method is based are considered to be part of the core genome and are not under diversifying selection (Enright et al. 2000; Feil et al. 2003). This has allowed both characterisation of *S. aureus* at the global species level as well as enabling the identification of the dominant lineages and their ancestral origins (Enright et al. 2000; 2002; Feil et al. 2003). Whilst there are large numbers of STs this work has also demonstrated that there are is a relatively small number of highly successful CCs worldwide, most notably 5, 8, 22, 30 and 45 (Feil et al. 2003). There are otherwise multiple globally distributed CCs including 1, 25, 59, and 121 (Feil et al. 2003; Nübel et al. 2011).

At the population level there are some notable features regarding CCs. From the initial MLST analysis it has been known that disease and carriage isolates are equally present across the major CCs, and are not associated with specific clonal backgrounds (Feil et al. 2003). Drug resistance on the other hand is associated with clonal background. A recent European surveillance study demonstrated that within CC5, 8 and 22 isolates studied more than 70% were methicillin resistant, whilst conversely methicillin sensitivity was universal in isolates of CC15 background (Aanensen et al. 2016). There are also examples of geographical distribution of certain CCs, for instance CC59 is highly prevalent in South East Asia (Chuang and Huang 2013), and similarly ST93 (derived from CC88) in Australia (Monecke et al. 2011).

1.1.2 The *S. aureus* genome

The initial studies of the *S. aureus* genome with both DNA microarray and subsequently whole genome sequencing (WGS) revealed that its contents could be classified into core and accessory regions (Fitzgerald et al. 2001; Lindsay and Holden 2004). The subsequent comparative analysis of these components revealed the dynamic nature of the genome, and the evolutionary mechanisms that have shaped it. These studies have led to major advancements in our understanding of how the organism both adapts to the host and environment, as well as to the emergence and spread of successful virulent and drug resistant lineages.

1.1.2.1 Conserved regions of the S. aureus genome

From the first genome studies of *S. aureus*, it became evident that a large proportion of the chromosome was conserved across the species. DNA microarray analysis of *S. aureus* across a wide array of CCs revealed that approximately 78% of the genetic material was conserved across the lineages (Fitzgerald et al. 2001). The first 5 sequenced *S. aureus* genomes represented globally derived hospital acquired methicillin resistant *S. aureus* (MRSA) (Kuroda et al. 2001; Holden et al. 2004), community acquired MRSA (Baba et al. 2002) and methicillin sensitive *S. aureus* (MSSA) isolates (Holden et al. 2004). Despite their diverse backgrounds comparison between these revealed that 75% of the predicted genes (ranging from 2592 to 2748) were conserved across all of these strains (Lindsay and Holden 2004). These core genes are typically involved in essential functions such as central metabolism and are shared across 95% of the species (Lindsay and Holden 2004). They also include genes essential for other species-wide survival functions such as virulence and host immune evasion (Lindsay and Holden 2006). Examples include toxins such as the membrane damaging α and σ haemolysins (*hla*/*hld*), surface binding proteins like fibronectin binding protein A (*fnbpA*), and proteases such as *sspA*.

Diversity in the core genome arises primarily through the accumulation of single nucleotide polymorphisms (Castillo-Ramírez et al. 2012), which accounts for diversity accumulated over short time scales (Lindsay and Holden 2006; Fitzgerald and Holden 2016). The impact of point mutations acquired in this region of the genome is both dependent on the type and its genomic position. Synonymous substitutions will have no obvious functional consequence as there is no resultant change in the amino acid sequence (Lindsay and Holden 2006). Conversely non-synonymous mutations, which change amino acid sequences, will have potential functional as well as phenotypic differences. The success of such alleles will then be dependent on whether they confer benefit or incur a fitness cost to the recipient in the context of their selective environment. For instance, gaining a single point mutation in the *agr* virulence regulatory system may provide short-term survival benefit in the context of invasive infection in the nosocomial setting, due to their ability to subvert host clearance as well as antimicrobial therapy (Fowler et al. 2004; Paulander et al. 2013). But the relative paucity of these mutations in clinical isolates generally is postulated to be because they result in a long term survival cost to the recipient (Shopsin et al. 2010).

1.1.2.2 Core genome variability between clonal complexes

Approximately 10% of the genome is composed of genes that vary between CCs but which are largely conserved within CCs (Lindsay et al. 2006), or alternatively, are only present within specific clonal complexes. Core variable (CV) genes are chromosomally encoded and therefore vertically inherited (Lindsay et al. 2006). Overall CV regions incorporate genes in functional classes ranging from regulatory systems, cell surface adhesive genes, and immune evasion genes (McCarthy and Lindsay 2010). These CV regions are believed to be contributory to observed lineage specific phenotypic variances.

The accessory global regulator operon contains four genes governing virulence regulation in accordance with cell density and growth phase. These include *agrA*/*C,* which form the two-component signal transduction system, and *agrB*/*D,* are involved in production of the autoinducer peptide (AIP) that then activates the system (Novick et al. 1995; Le and Otto 2015). Whilst all four genes are conserved across the species four variants exist (types I-IV), and fall within specific clonal complexes. These are thought to have arisen ancestrally through horizontal gene transfer (Robinson et al. 2005).

Each *S. aureus* lineage possesses a varying complement of surface proteins (Sung et al. 2008). Some examples of these genes are highly conserved across the species such as *Spa*, *isdA*, *and fnbpA* for instance. Others are have been shown to be sporadically absent such as *clfA*, *clfB* and *sdrD* (McCarthy and Lindsay 2010). Recombination is thought to have shaped the lineage variable presence of some of these surface genes. Examples of this being the absence of a distinct *fnbpB* allele in CC22 and 30 background (Holden et al. 2013), or lack of *sdrD* in CC30 (Lindsay and Holden 2006). There are also examples of the adhesive genes that are completely absent from major *S. aureus* lineages, such as the collagen adhesins absence in CC5 and 8, or the absence of surface protein G (*sasG*) from CC30 and 45 (McCarthy and Lindsay 2010).

Immune evasion determinants are widely distributed throughout the *S. aureus* genome, with many of them conferred by carriage on mobile genetic elements (MGE). The genomic islands (GI) α and β are believed to have arisen through horizontal gene transfer and continue to evolve. These regions are found in all reference genomes, however vary extensively in their constituent components, in the case of GIα (McCarthy and Lindsay 2013). Whilst in GIβ they vary on the basis of the lineages which they are present in (McCarthy and Lindsay 2013). The GIα locus for instance encodes multiple superantigen like toxins (*ssl*) that have wide ranging functions and is one of the most variable immune evasion loci (Lindsay and Holden 2004;

McCarthy and Lindsay 2013). There are both differing numbers of the *ssl* genes within different genetic backgrounds as well as variable numbers of allelic variants of each *ssl* gene. For example *ssl3* is noted to have up to 13 allelic variants (McCarthy and Lindsay 2013).

1.1.2.3 The Accessory genome

The accessory genome is composed of mobile genetic elements (MGEs), which have entered the genome horizontally, and accounts for approximately 15% of the chromosome (Lindsay 2014). It includes bacteriophage, plasmids, transposons, *S. aureus* pathogenicity islands (SaPIs) and Staphylococcal cassette chromosomes (SCC). These components often carry clinically relevant virulence and resistance determinants. MGEs can therefore contribute to different phenotypes observed between isolates of the same clonal background. For instance the reference isolates MSSA476 and MW2 are both ST1, and differ by relatively small core genome SNP differences of 285 bp in orthologous genes despite being isolated in different continents. Yet they differ in their antimicrobial susceptibility and toxins profiles because of their accessory genome content (Holden et al. 2004). The prime example in this instance is carriage of the Panton-Valentine Leukocidin in MW2 which is associated with skin and soft tissue infection (Baba et al. 2002).

Horizontal gene transfer (HGT) arises in *S. aureus* by three mechanisms which include transduction, transformation and conjugation (Lindsay 2014). The bacteriophage (converting phage) can infect and introduce virulence factors that are responsible for wide ranging disease phenotypes caused by the organism. These include the following very specific examples: exfoliative toxin A (*eta*) responsible for Staphylococcal Scalded Skin syndrome (Amagai et al. 2000), enterotoxins such as enterotoxin A (Betley and Mekalanos 1985) which are responsible for emesis in food poisoning, Panton-Valentine Leukocidins lukS/F which are associated with severe skin and soft tissue infection and pneumonia (Kaneko et al. 1998). Chromosomally

integrated bacteriophage, known as prophage, can be induced under conditions of stress such as during antibiotic exposure leading to their excision from the chromosome and transfer onto to other cells. This is considered as a mechanism of adaptation to a host or environment, and has been shown to occur in both asymptomatic carriage and infection (McCarthy et al. 2012).

The SaPIs are another component of the accessory genome encoding important virulence determinants including examples such as the superantigens enterotoxins B, C, and K as well as the Toxic Shock Toxin (tst) (Lindsay and Holden 2006). These accessory components are transferred horizontally by helper phage (Lindsay et al. 1998). There is a lineage specific distribution of SaPIs, an example being the SaPIs carrying *tst* are found in a limited number of lineages such as CC30 and 45 (Moore and Lindsay 2001; Holtfreter et al. 2007).

This accessory genome is resident to the MGEs conferring antimicrobial resistance. 3 broad classes of MGE carry these determinants including: SCC, plasmids and transposons. The SCCs will be discussed in detail in section 1.3. The plasmids present in *S. aureus* are generally categorised on the basis of size and also whether their transfer is mediated by conjugation. Class I plasmids are the smallest, approximately 5kb, and are primarily transferred by transduction. These carry resistance determinants such as to bleomycin or kanamycin as seen in pUB110 (Lindsay and Holden 2006). Class II plasmids, such as pSAS are larger, 40kb or less, and carry resistance determinants to beta-lactams and heavy metals like cadmium, and are also transferred by transduction (Lindsay and Holden 2006). Class III plasmids are generally 45kb or more, carry a variety of resistance determinants including to heavy metals such as arsenic, and have transfer genes (*tra*) which mediate their transfer by conjugation (McCarthy and Lindsay 2012). Transposons frequently encode resistance determinants, with the most obvious examples being Tn*554*, which confers erythromycin resistance via *ermA* (Phillips and Novick 1979). Other pertinent examples include Tn*552* and Tn*5801* which carry beta-lactamase *blaZ* and tetracycline resistance *tetM* determinants respectively (Rowland and Dyke 1989;

Kuroda et al. 2001). Their transfer can be mediated by several mechanisms, either by excisionary function of their encoded transposase gene, by insertion into another MGE such as a plasmid, or rarely by conjugation such as has been described in Tn*916* (Franke and Clewell 1985).

1.1.2.4 Recombination and its impact on the genome

Variation in the core genome, in addition to the accumulation of SNPs, can also arise through homologous recombination. Whilst SNPs arise frequently, recombination is thought to contribute to genomic variability to a lesser extent. Using MLST it has been estimated that point mutations contribute to core genome variability with a 15 fold greater frequency than recombinatory events (Feil et al. 2003). This is in direct contrast to panmictic bacterial species, which are considered highly recombinogenic such as *Neisseria meningitidis* (Feil et al. 1999) or *Helicobacter pylori* (Suerbaum et al. 1998). Nonetheless, several *S. aureus* core genomic regions are thought to have been shaped by recombination including the coagulase (*coa*) and accessory global regulatory loci (*agr*) (Watanabe et al. 2009).

Whole genome sequencing (WGS) studies have revealed regions within the genome where recombination occurs contributing to diversification of the species. In a comparative study of globally derived ST239 MRSA isolates, it was shown that recombination was restricted to accessory regions of the genome (Castillo-Ramírez et al. 2012). Recombination rates within core genome were in keeping with previous estimates derived from MLST studies, overall indicating that recombination does not extensively contribute to genomic diversification over timescales of decades (Castillo-Ramírez et al. 2012). More recently, comparison of genomes from multiple strain backgrounds across the species demonstrated that regions susceptible to recombination included those adjacent to MGEs and in proximity to the origin of replication (Everitt et al. 2014).The findings of this work revealed that at the species level there was

evidence of widespread core genome recombination, but little evidence of this arising in closely related strains (Everitt et al. 2014). The authors concluded that recombination therefore contributes to the diversity across the species rather that at strain level (Everitt et al. 2014).

In contrast to these overall gradual recombination events across the species, there is evidence of single dramatic events that can shape lineages. The chimeric strain ST239 exemplifies the impact of recombination leading to the birth of a new successful lineage. In this instance a large DNA replacement led to the a hybrid chromosome formed from an ST8 strain, but with a 635kb fragment derived from a CC30 donor strain (Holden et al. 2010).

1.1.3 Niche adaptation

Key to the success of this organism is its ability to transmit and survive on a pandemic scale, and within wide ranging ecological niches. Examination of the genome has revealed its plasticity and capacity to adapt to new hosts and environments.

1.1.3.1 Host

Testament to this organism's adaptability is its versatility with regards to the host environments it can inhabit. As with typing studies of human isolates, molecular typing methods have revealed evidence of successful lineages correlating to specific animal hosts.

Early studies of animal isolates revealed evidence of association between CC97 isolates in bovine hosts and CC5 isolates in avian hosts (Smith et al. 2005; Smyth et al. 2009) for instance. Subsequent WGS of isolates from these hosts have shown how the organism has undergone specific genetic adaptations to survive in that niche host, as well as adaptations that allow the jump from human to animal carriage. The pig associated CC398 MRSA is an exemplar. In its

transition between human and livestock associated carriage there was loss of the immune evasion complex, and of both methicillin and tetracycline resistance (Price et al. 2012). The loss of the prophage, carrying *scin*/ *sak*/*chips* genes demonstrates an example of loss of an MGE that presumably only conferred benefit to evasion in the human host given its reported general absence in animal isolates (van Wamel et al. 2006; McCarthy and Lindsay 2013). Whilst acquiring novel MGEs can aid the success of a lineage in a new host much finer genetic change has been shown to allow such a host switch. In the CC121 lineage, which had been associated with skin infections in humans, a single SNP in the cell membrane protein *dltB,* has been shown to have led to the lineages jump from human to rabbit hosts (Viana et al. 2015).

1.1.3.2 Environment

From their earliest emergence methicillin resistant *S. aureus* lineages have been prevalent in the hospital environment. Each of the early endemic MRSA clones; including ST5, ST239 and ST36 shared the common feature of being multi-drug resistant in addition to being methicillin resistant (Uhlemann et al. 2014b). With each there was also a carriage of variants of the SCC*mec*, within which the methicillin resistance determinant *mecA* was found. In the early 1990s incidences of community acquired MRSA (CA-MRSA) began to be reported in previously healthy individuals with no risk factors for contracting hospital-acquired MRSA (HA-MRSA) (DeLeo et al. 2010). Whilst initially reported in remote regions of Australia similar cases began to emerge from the USA of severe disease being caused by MRSA acquired in the community (Udo et al. 1993; CDC 1999). Subsequent studies revealed that the newly emerged CA-MRSAs had several unique features compared to their HA counterparts. These included susceptibility to non-β lactam antibiotics, PVL (Panton-Valentine Leukocidin) positivity, strong phenotypic association with skin/soft tissue infections, and their ability to cause fatal disease in healthy children and adults (DeLeo et al. 2010). Analysis of the genomes of two examples of these virulent community strains revealed that they had a type IV SCC*mec* element (Baba et al. 2002;

Diep et al. 2006), which was notably smaller than the type II SCC*mec* found in the ST5 MRSA N315 (Kuroda et al. 2001). This is believed to be one factor underpinning the successful spread of CA-MRSA and as well as the HA-EMRA15 lineage (Knight et al. 2012).

In CA-MRSA, carriage of other virulence determinants and not resistance genes are believed to contribute to their rapid transmission and pathogenicity. The highly successful CC8 clone, USA300, carries an arginine catabolic mobile element (ACME) which is thought to be contributory to its spread (Diep et al. 2006). The arginine deaminase present within this element is proposed to provide a mechanism to allow *S. aureus* to survive the acidic pH of the skin encountered during colonisation (Thurlow et al. 2013). The increased virulence of USA300 has also been explained by lineage specific differences in the expression of haemolysin α, which is another important difference between hospital and community clones (Li et al. 2009b).

1.2 Variation occurring in the genome over short time scales

1.2.1 Within host diversity during disease

In 2007 the first study examining the host evolution of *S. aureus* during invasive infection was published (Mwangi et al. 2007). This seminal study used whole genome shotgun sequencing of sequential bacteraemia isolates from a single patient to demonstrate the accumulation of genome wide mutations. In total 35 unique SNPs were identified in 9 isolates spanning 3 months in an individual receiving broad-spectrum antibiotics (Mwangi et al. 2007). This analysis revealed time synchronous accumulation of mutations conferring resistance to the antimicrobials being used to treat the patient, including rifampicin and vancomycin. The sequential mutations matched the observed increasing phenotypic resistance to these antibiotics, as well as development of daptomycin resistance despite it not having been used in

the patient (Mwangi et al. 2007). This study therefore provided the first insights of within host diversity and evidenced bacterial adaptation driven by clinical interventions in the patient.

These findings illustrated the micro-evolution of core genome SNPs within a clinically relevant time frame. From a genomic perspective, the mutation rate came from population level studies using WGS to define the diversity of globally derived ST239 isolates (Harris et al. 2010). This work revealed evidence of convergent evolution of antimicrobial resistance across the lineage, and specific geographical sub-lineages correlating to the country of origin. For the first time it allowed calculation of genome wide mutations rates for this organism (Harris et al. 2010) meaning that times scales for micro-evolutionary change could be defined. It was also the first to demonstrate that samples from patients in a single hospital could be linked on the basis of short genetic distances between the isolates thus demonstrating putative transmission events (Harris et al. 2010). Following on from this mutation rates were calculated from studies of other pandemic MRSA lineages, revealing that the genome of this organism diversifies at different rates depending on its genetic background. Rates have been estimated at between 1.2 $x10^{-6}$ substitutions per base per year in USA300 to 2.0 $x10^{-6}$ in ST225 (Uhlemann et al. 2014a), but as high as 3.3 $x10^{-6}$ in the chimeric ST239 lineage (Harris et al. 2010). These advancements have meant that the observed diversities across populations of isolates, and in individuals can be used to date the time to most recent common ancestor. This has provided the ability to date the emergence of successful lineages on the basis of the observed genetic heterogeneity within a clonal collection of isolates (Holden et al. 2013).

1.2.2 Between host variability

Characterising diversity in populations of closely related isolates has been a successful strategy for tracing MRSA transmission in clinical settings where traditional typing methods and epidemiology would not have allowed source identification. One of the first studies to demonstrate this was undertaken in a neonatal intensive care unit, where outbreak isolates were linked using WGS because of their relative SNP distances (Köser et al. 2012). This study also utilised common toxin and resistance profiles in the isolates to correctly rule in and out of the outbreak groups (Köser et al. 2012). In the wake of this WGS was then used to trace transmission networks and the source in a prolonged MRSA outbreak in a special care baby unit (Harris et al. 2013). Deep sequencing of isolates from a single member of staff in this study revealed that they carried "a cloud of diversity". This referred to the genetic heterogeneity that had accumulated in this individual's clonal colonising population during what was shown to be an extensive period of asymptomatic colonisation (Harris et al. 2013). This deep sampling allowed the identification that one individual shared part of their colonising population with multiple individuals spanning the whole outbreak period (over 6 months). By studying the phylogeny of 20 sequenced colonies from one individual, they were able to show these populations were genetically closely related to those found in infants affected by the outbreak over temporally separated periods (Harris et al. 2013).

1.2.3 Bacterial Adaptation to the host

Diversity studies have revealed dynamic changes in bacterial genomes over short time scales (days to months). They have shown the influence of host factors such as exposure to antibiotics (Mwangi et al. 2007) in driving bacterial adaption to survive in the hosts. Studies that have characterised bacterial phenotypic and genotypic changes over more extended

periods (upto decades) have provided unique insights as to how pathogens evolve to adapt to their host. There are interesting examples where pathogens of very significant disease potential can down regulate their virulence to persist quiescently in a host. For instance *Neisseria meningitidis*, a causative organism of bacterial meningitis, can be asymptomatically carried in the nasopharynx for extended periods of time. This organism has been demonstrated to down-regulate expression of two genes encoding surface proteins (*fetA/ nadA*) during colonisation of persistent carriers (Alamro et al. 2014). This is believed to be a mechanism by which the organism avoids detection and clearance by the host. Another example is the highly pathogenic bacterium *Burkholderia pseudomallei*, which causes the commonly fatal disease Melioidosis. A study comparing the genome sequences of two isolates of *B. pseudomallei* from a single patient isolated 12 years 7 months apart, revealed an accumulation of mutations in capsular polysaccharide and lipopolysaccharide loci that were believed to have accompanied the organism switching from highly pathogenic to a rarely observed commensal-like state (Price et al. 2013).

Perhaps the best example of bacterial adaptation can be found in studies of Cystic Fibrosis (CF), the chronic bronchiectatic lung disease characterised by recurring infective exacerbations. It is associated with long-term carriage (up to decades) of the environmental pathogen *Pseudomonas aeruginosa*. Genotypic and phenotypic studies of bacterial isolates from patients with CF have shown global adaptive mechanisms used by the organism to adapt to the host environment.

Extensive genomic research in *P. aeruginosa* in CF has identified evidence of convergent evolution both within and across study populations, providing insight of important pathogenic mechanisms in the disease. A longitudinal study of 34 children and adults with CF revealed 52 mutated genes across the more than 400 samples sequenced from the sampled population (Marvig et al. 2015). Of these homoplastic mutations, multiple examples were in specific loci

previously identified as being mutated in chronic carriage of *P. aeruginosa* in CF including, genes associated with antimicrobial resistance and biofilm formation. They also demonstrated that non-synonymous mutations in transcriptional regulators, cell capsule and toxin genes were significantly over-represented in the homoplasies identified, indicating their importance in the organisms pathogenicity (Marvig et al. 2015). The identification of mutations in quorum sensing regulator *lasR*, involved in virulence regulation, by this study also conformed with findings from a previous longitudinal study indicating the trend towards reduced virulence during persistence (Smith et al. 2006; Marvig et al. 2015).

Mutations in DNA mismatch repair genes such as *mutS* and *mutL* which lead to hypermutability, are another adaptive trait that has been identified in both longitudinal and individual patient population studies (Smith et al. 2006; Marvig et al. 2013). This is seen as a mechanism whereby there is accelerated supply of novel genetic variants that may confer survival advantage in the host (Moxon et al. 2006; Didelot et al. 2016). Whilst this mechanism would provide an obvious and rapid supply of genetic variants subtler mutations have been shown to confer adaptive capacity. Having identified an accumulation of mutations in the promoter regions of the iron scavenging *phu* system across a patient population (Marvig et al. 2013; 2014), the authors confirmed *in vitro* that this mutation aided survival of *P. aeruginosa* by allowing it to scavenge haemoglobin from the host (Marvig et al. 2014).

In addition to these clear signals of adaptation at population level, it has also been shown that *P. aeruginosa* can adapt to specific niche sites within an individual host. Examination of samples from a single CF patient derived across a 32-year period showed that a single colonising strain had formed 3 distinct “sub lineages” in each lung and the sinuses independently (Markussen et al. 2014). These sub-populations were then shown to have varying genotypic and phenotypic diversity, with distinct mutational rates.

1.2.3.1 S. aureus adaptation to the host during infection

The findings in *P. aeruginosa* demonstrate the overall trends of adaptive strategies in long-term host pathogen interaction. These adaptations reflect the significant change the organism must undergo in its transition from an environmental organism to the unique host environment of the CF lung. Increasing antimicrobial resistance and declining virulence are the clearest examples of adaptive mechanisms allowing an organism to persist within the host. The adaptive potential of *S. aureus* is evidenced by its ability to infect and persist in such a wide variety of host settings, ranging from the skin and soft tissue, to bones and joints and even within the bloodstream. Whilst a limited number of studies have utilised WGS of *S. aureus* isolates to identify adaptation to the host, there are clear examples of adaptive mechanisms, which mirror those identified in studies of *P. aeruginosa.*

Studies employing WGS of multiple and/ or sequential *S. aureus* colonies from individuals during nasal carriage and disease have increased our understanding of within host evolution of this organism and the factors influencing it. As shown in Figure 1-1, this process begins with acquisition of the organism, which is typically by colonisation. As the population expands on the host it is subject to selective pressures such as therapeutic interventions or host immune response (Didelot et al. 2016). The potential outcomes in the host are as follows: clearance, onward transmission to another host, infection, or adaptive changes that have the potential to confer a survival benefit in the host. Persisting populations can then remain inertly, or subsequently result in infection within the individual (Didelot et al. 2016). Where onward transmission occurs there is donation of the diversity that has accumulated during the time in the donor. The following summarises evidence of in host adaptation specifically in *S. aureus*.

1.2.3.2 Evidence of selective pressure of antibiotics in the host

The evolution of antimicrobial resistance during antimicrobial therapy is the most widely studied example of the impact of selective pressure in the host. Both in *S. aureus* and across bacterial species generally, multiple core genes have been shown to accrue mutations in direct response to antibacterial therapy. Of these adaptations, multiple studies have demonstrated mutations in the cell wall metabolism genes *vraR* and *walkR* conferring vancomycin resistance (Mwangi et al. 2007; Howden et al. 2011; van Hal et al. 2014), RNA polymerase *rpoB* leading to rifampicin resistance (Mwangi et al. 2007; Dordel et al. 2014) and DNA gyrase subunit A (*gyrA*) leading to fluoroquinolone resistance (Azarian et al. 2016). Similar findings have been shown in response to the use of Linezolid, generally reserved for MRSA infection, where point mutations in ribosomal methyltransferase *rlmN* led to reduced susceptibility and treatment failure (Gao et al. 2010). In another example, development of reduced susceptibility to the third line antibiotic daptomycin has been shown to arise due to mutations in the phospholipid synthesis genes *cls2* and *pgsA* (Peleg et al. 2012).

1.2.3.3 Regulatory reprogramming through mutation

Attenuating overall virulence to the host is seen as another key adaptive strategy employed by *S. aureus*. The importance of the *agr* system, which regulates the expression of toxins and adhesins in response to cell density, is clear from experimental evidence. In animal infection models *agr* mutants show markedly reduced disease severity (Cheung et al. 1994; 2011). Whilst clinical evidence has shown *agr* dysfunction to be associated with persistence during bacteraemia and poor clinical outcomes (Fowler et al. 2004).

Multiple studies have reported *agr* mutations during carriage or disease. For instance two separate loss of function *agrA* mutations were shown in the veterinary hospital study during the course of sampling of a progressive invasive MRSA infection in a dog host (Paterson et al. 2015). A similar example was reported in an individual with a chronically MRSA-infected foot ulcer (Azarian et al. 2016). Interestingly mutations in *agrC* have been reported in association with the development of vancomycin resistance during prolonged antibiotic exposure (Mwangi et al. 2007; Chen et al. 2014), although the significance of this is unclear. Agr mutants are thought to have a survival advantage in antibiotic selective environments (Paulander et al. 2013), which is thought to be one of the reasons they are observed at much higher frequency in nosocomial (hospital) populations in comparison to community isolates (Shopsin et al. 2008; 2010).

Mutations in other regulatory systems are an alternative route for the organism to reduce its virulence during prolonged interaction with a host. The longitudinal follow up of a single nasal carrier who progressed to fatal bacteraemia led to the identification of a truncating mutation in an AraC-family transcriptional regulator (*rsp*) during the conversion to invasive disease (Young et al. 2012). This gene has subsequently been shown to have control in toxin production as well as expression of immune evasion proteins (Das et al. 2016), which when mutated reduced virulence but without impacting on the ability to cause invasive infection. Interestingly a non-synonymous mutation in this gene family has also recently been shown in an isolate from a chronic abscess in a diabetic patient with persistent USA300 infection (Azarian et al. 2016).

1.2.3.4 Phenotypic variation in host

Phenotypic switching is another adaptive strategy employed by *S. aureus* to survive the host environment. Small colony variants (SCV) are an example of this whereby the organism slows

its metabolic rate, and forms colonies around 10% of the size of normal colonies (Proctor et al. 2006a). SCVs are associated with globally reduced virulence, persistent infection, increased antimicrobial resistance and intracellular growth (Sendi and Proctor 2009). From a genetic perspective several mutations are known to confer SCV phenotype, including mutations in the electron transport chain *hemB* (Balwit et al. 1994), and *fusA* (translation elongation factor). (Norström et al. 2007). The biochemical basis for SCVs is auxotrophism for menandione, thymidine or haemin causing reduced tricarboxylic acid cycle metabolism (Proctor et al. 2006b). This phenotype is not fixed during chronic infection but is dynamic, and can be readily reversed during nutrient rich conditions (Tuchscherr et al. 2011) . This demonstrates that it is both an adaptation to subvert host clearance but also to survive environmental stress.

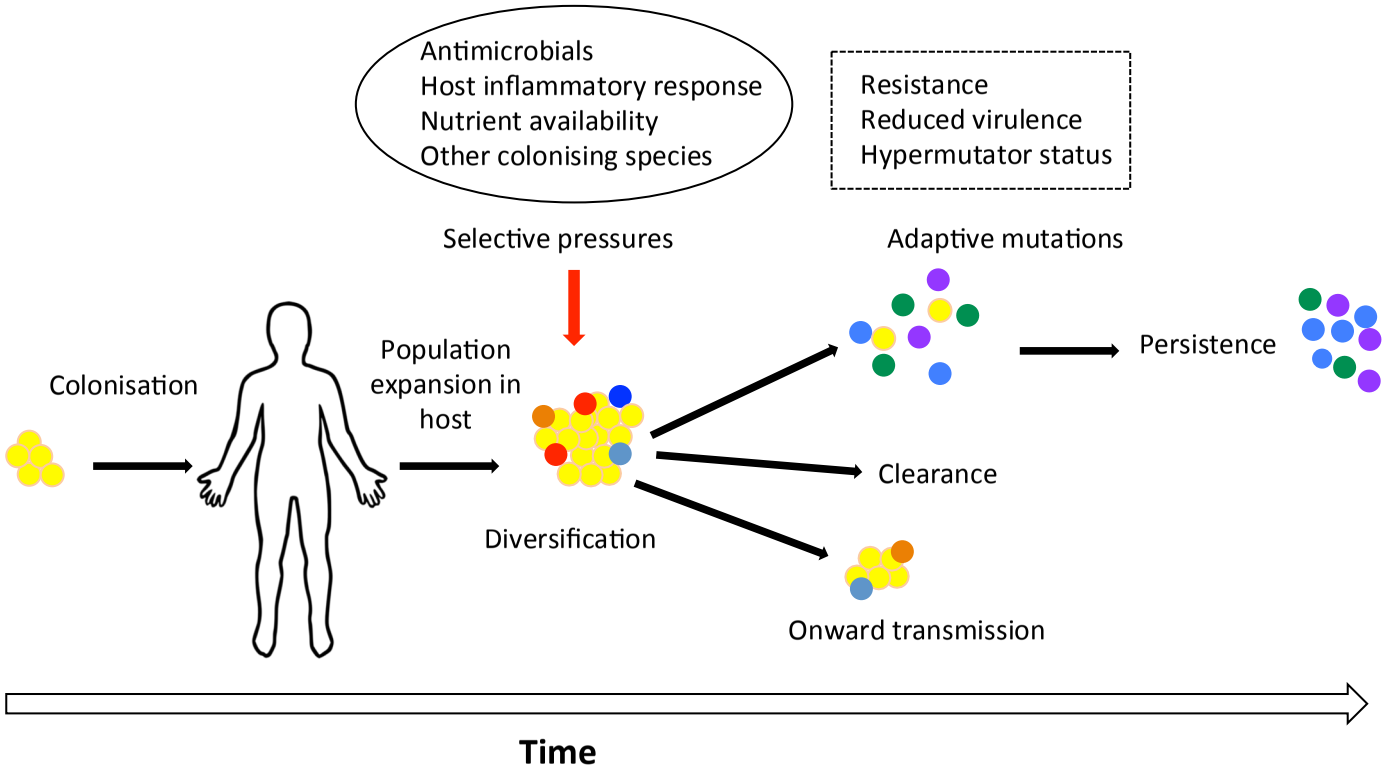

Figure 1-1 In host diversity and adaptation in S. aureus. Colonisation begins with S. aureus (yellow circles) acquisition from an external source. Population expansion with accumulation of core genome mutations (diversity) in coloured circles. Selective pressures that the colonising population are exposed to in host (examples within oval) drive adaptive mutations (examples in rectangle), or lead to clearance. Population can during this period can also be transmitted on to other sources with donation of population diversity (purple/ green and blue circles). Favourable adaptive mutations allow for continuing persistence in the host. Persistor populations can then remain as a carrier state or subsequently lead to infection in the host.

1.2.3.5 S. aureus adaptation to colonisation

Asymptomatic carriage of *S. aureus* occurs with much greater frequency than infection. Despite this little is currently understood of how the organism potentially adapts during this more inert interaction with the host and the unique ecological niches presented by different individuals.

1.2.3.5.1 The blurred boundaries of S. aureus interaction with the host

The terms colonisation, carriage and commensal are frequently used, but there is a lack of clarity between and across studies to what state each is actually addressing in terms of bacterial interaction with the host. One definition of colonisation is a state during which there is an interaction between a micro-organism and a host, which gives rise to a variable amount of host damage, ranging from minimal to great. Included in this definition is the need to take into consideration the host immune response, and its ability to clear the organism before damage arises (Casadevall and Pirofski 2003). Commensalism alternatively can be considered as a host-microbe interaction which does not result in host damage after acquisition (Casadevall and Pirofski 2003). These definitions are important to consider, as carriage is presumptive of colonisation. From these definitions colonisation can result in damage which is t as perturbations in the normal mechanisms that inhibit host entry or the organism's ability to establish itself in the host (Casadevall and Pirofski 2003).

Nasal carriage studies are therefore probably representative of both commensal and colonisation populations. From a species level it has been proposed that *S. aureus* is already adapted to human carriage because of its long-term evolution with humankind (van Belkum et al. 2009a). The genotypes of the dominant *S. aureus* lineages

therefore potentially already reflect their adaptation to the human host, and primarily the nasopharynx, the commonest site for isolation of this organism. Humans as a species are not uniform either genetically or phenotypically (1000 Genomes Project Consortium et al. 2010). Each individual represents a different ecological environment for the organism to inhabit, varying in age, sex, health status as well as in their microbiome composition (Grice et al. 2009; van Belkum et al. 2009a). Wide ranging host factors are proposed to account for the variable carriage statuses, and especially those termed as persistent carriers (van Belkum et al. 2009b). Host states associated with higher rates of carriage include for instance diabetes mellitus, chronic renal failure (dialysis-treated), human immunodeficiency virus (HIV), and the extremes of age (van Belkum et al. 2009a). There are also examples of host genetic polymorphisms proposed to be associated with increased carriage such as glucocorticoid receptor gene polymorphisms (van den Akker et al. 2006). The microbiome and its variation between individuals is also very likely to influence this process given what is known about co-species antagonism during colonisation between *S. epidermidis* and *S. aureus* as one illustration (Iwase et al. 2010; Lai et al. 2010). This would suggest that the organism must also adapt during asymptomatic carriage to its specific hosts' niche.

1.2.3.6 Current limited evidence of adaptive changes during nasal carriage

Few studies have specifically assessed *S. aureus* asymptomatic carriage populations for evidence of adaptive evolution during the period of colonisation. The study carried out by (Golubchik et al. 2013) undertook sequencing of isolates from 13 community nasal carriers to look for evidence of convergent evolution during carriage. They identified potential adaptive changes including mutations in surface anchored proteins SasC and ebh and in a single enterotoxin gene (Golubchik et al. 2013). Whilst in long-term carriers samples by (Young et al. 2012) there were no reported homoplastic mutations across the 3 individuals sampled, there were individuals whose populations similarly contained non-synonymous mutations in surface

proteins sasC and sasG for instance. Transmission studies have provided a glimpse in host diversity arising during carriage. Comparison of the mutations identified by deep sampling in such studies demonstrates that there are similar gene classes reportedly mutated in these study populations, for instance mutations that resulted in truncation in surface proteins Spa and SdrC/D/E found in long-term MRSA carriers in a vet hospital (Paterson et al. 2015). Given these proteins function in attachment to host surfaces it could be postulated that this would enhance transmissibility by virtue of reducing adhesion to the primary host. It has also been proposed that mutations in surface proteins are observed in such studies, as they are antigens directly interacting with the host. As such they are under diversifying selection allowing epitopes that will subvert host immune detection to arise (Didelot et al. 2016).

Comparing mutations reported in carriage, transmission and invasive disease isolates reveals a limited number arising in shared loci or gene families that are suggestive of adaptive evolution. As a single example one long-term nasal carrier in the study by Paterson et al (2015) had a total of 84 SNPs, 13 deletions and 5 insertions present within the individuals colonising population. As part of this heterogeneity were mutations in *agrC* and the *AraC* transcriptional regulator (Paterson et al. 2015) which exemplify the gene classes shown during in host evolution of both active invasive or transition to invasive disease (Mwangi et al. 2007; Young et al. 2012; Azarian et al. 2016). Whether these are therefore representative of adaptive change to colonisation cannot be concluded. As proposed by (Golubchik et al. 2013) clonal carriage populations will naturally expand and diversify, then contract during clearance. Longitudinal sampling over this cycle would be the obvious route to identifying genetic diversity favouring survival during this contraction period, and hence specific adaptive changes, which are currently very limited.

1.3 The evolution of antimicrobial resistance in S. aureus

The strongest evidence of selection shown within patients is the development of resistance during antibacterial therapy. At a population level and over much greater timescales, arguably the most significant adaptation that *S. aureus* has undergone from mankind's perspective is the development of drug resistance.

Multi-drug resistant populations of this pathogen population are now a global health concern. In his 1945 Nobel Prize lecture Sir Alexander Fleming highlighted the dangers of widespread availability of the antibiotic Penicillin, and how inappropriate dosing during treatment would readily select for microbial resistance (Fleming 1945). By the time of acceptance of this Nobel prize Penicillin resistance was being already being reported in nosocomial *S. aureus* isolates (Kirby 1944; Bondi and Dietz 1945; Spink and Ferris 1945).

1.3.1 Timeline of emerging antimicrobial resistance in *S. aureus*

With each successive class of antibiotic introduced into clinical practice *S. aureus* has developed mechanisms to evade their action, and generally within only a few years of the drug being introduced (Figure 1-2). This began with the introduction of penicillin in the 1940s progressing to multi-drug resistant strains that have acquired mechanisms to evade vancomycin by the late 1990s (Chambers and DeLeo 2009a).

The sulfa-drugs were the first antibiotic class used in humans (Silver 2011; Lewis 2013). Of these Prontosil was the first agent tested in humans (Domagk 1936). Developed as a red dye by the Bayer laboratories in Germany, subsequent screening for the antimicrobial properties of dyes revealed its activity was against haemolytic Streptococci (Domagk 1936). Its use

marked the advent of the so-called "golden era of antibiotics" which spanned the 1940s to 60s. It was during this period that the majority of the most widely used antimicrobial classes were either discovered and or introduced into clinical practice. In this interval naturally derived compounds from soil-based organisms were some of the most successfully introduced, with Penicillin being the primary and most famous example. However even before widespread use of antibiotics for *S. aureus* infection and more generally, resistance had already been identified *in vitro* against penicillin and the earliest incarnation of the tetracyclines, Aureomycin (chlortetracycline) (Abraham and Chain 1940; Demerec 1949) (Figure 1-2).

Streptomycin and Chloramphenicol were two of the earliest agents introduced into clinical practice in the so-called golden era. Although their initial usage was not targeted against *S. aureus* nonetheless resistance emerged (Figure 1-2). The exact date of the first instances of resistance to these drugs in *S. aureus* are not clear, which may reflect that this was not directly assessed for during the first years of usage for *Mycobacterium tuberculosis* (Pyle 1947) and *Salmonella typhi* (Woodward and Smadel 1948) respectively. Whether the subsequent resistance that emerged in *S. aureus* related to its use against that organism, or as an indirect effect on it as a commensal, is also not possible to determine.

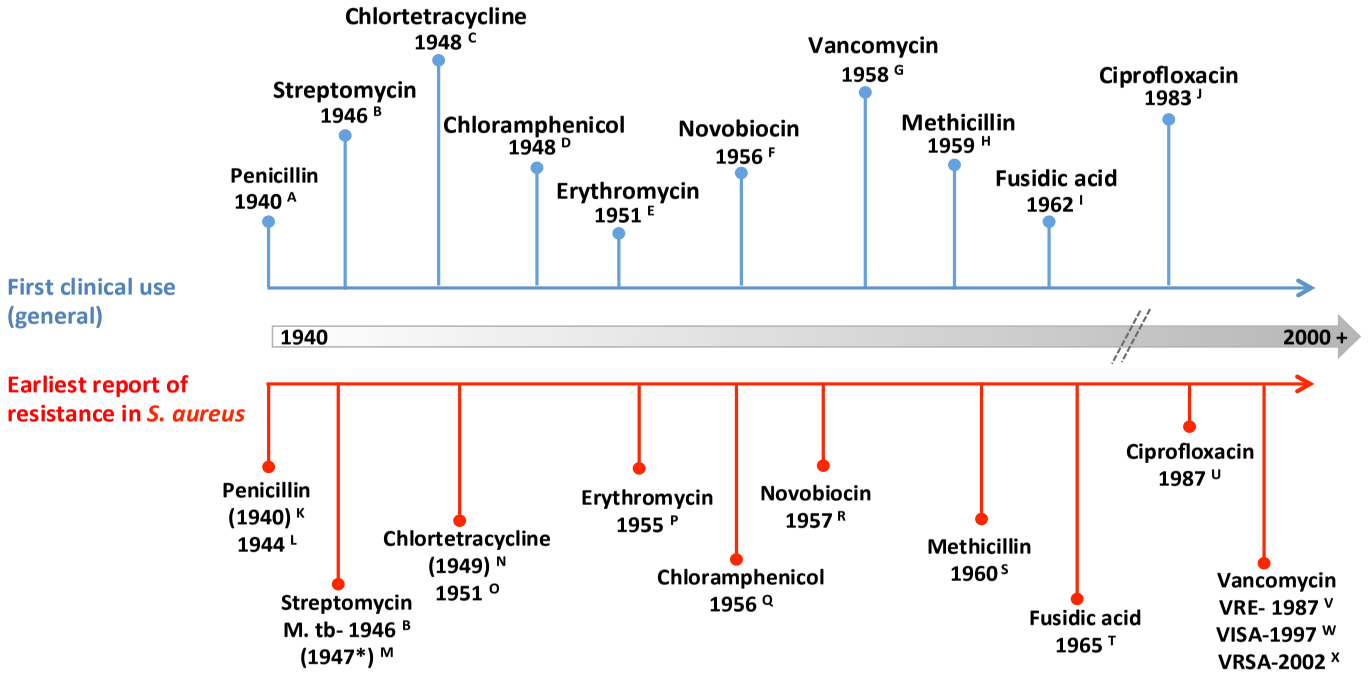

Figure 1-2 Antimicrobial introduction from 1940 onwards and the emergence of resistance in *S. aureus*. Examples of the major classes of antibiotic introduced into clinical practice between 1940s to 1980s (above timeline). Earliest clinical reports of resistance in *S. aureus* are indicated below the timeline. Dates in brackets indicate that the first identification of resistance was initially from *in vitro* analysis (Penicillin and Chlortetracycline). Date in brackets with asterisk indicates that first detection of resistance was during *in vitro* by serial passage (Streptomycin). VRE- Vancomycin resistance enterococci, VISA- vancomycin intermediate sensitivity *S. aureus*, VRSA- vancomycin resistant *S. aureus*. **References:** **Penicillin (A)** (Chain et al. 1940) **(K)** (Abraham and Chain 1940) **(L)** (Kirby 1944); **Streptomycin (B)** (Pyle 1947) (M)(Klein and Kimmelman 1947); **Chlotetracycline (C)** (Schoenbach et al. 1948a) **(N)** (Demerec 1949) **(O)** (Levinson et al. 1951); **Chloramphenicol (D)** (Woodward and Smadel 1948) **Erythromycin (E)** 8(McGuire et al. 1952) **(Q)** Roantree and Rantz 1955)); **Vancomycin (G)** (Levine 2006) **(V)** (Leclercq et al. 1988) **(W)** (Hiramatsu et al. 1997) (Chang et al. 2003); **Methicillin (H)** (Knox 1960) **(S)** (Jevons 1961); **Fusidic acid (I)** (Newman et al. 1962) **(T)** (Evans and Waterworth 1966); **Ciprofloxacin (J)** (Crump et al. 1983) **(U)** (Mulligan et al. 1987; Peterson et al. 1990).

Discovered in 1948 (Duggar 1948), the tetracycline aureomycin (also known as chlortetracycline), was shown to induce resistance in *vitro* (Demerec 1949), and shortly thereafter demonstrated *in vivo* resistance developing during treatment of a case of bacterial endocarditis (Schoenbach et al. 1948; Levinson et al. 1951). Similar findings were then reported with the macrolide erythromycin (known initially as Ilocytin), which following its discovery in 1951 (McGuire et al. 1952) was found to induce resistance leading to treatment failure in a patient with endocarditis (Roantree and Rantz 1955). Following this Novobioicin, an ATPase inhibitor of DNA gyrase and topoisomerase IV (Bisacchi and Manchester 2014), was introduced as an anti-staphylococcal agent in 1956. It followed the same pattern with spontaneous development of resistance during treatment reported a year later (Fairbrother and Williams 1956; Colville et al. 1957).

The first approval of Vancomycin for clinical usage came in 1958 in the USA, however it was not widely used at this time as methicillin received similar approvals at the same time (Levine 2006). Its potential toxicity from initial trials meant that methicillin was used as the first line therapy for resistant staphylococcal infections. In the following decades the increasing burden of MRSA led to a 100 fold increase in the use of Vancomycin by the 1980s (Kirst et al. 1998) Following on from this the first Vancomycin resistance in *Enterococcus faecium* was identified in 1987 (Leclercq et al. 1988). The first reports of vancomycin intermediate sensitivity *S. aureus* (VISA) (MIC 8-16 μg/ml) and vancomycin resistant *S. aureus* (MIC > 32 μg/ml) (VRSA) then emerged in 1997 and 2002 (Hiramatsu et al. 1997; Chang et al. 2003).

One the last drugs introduced in this era was Fusidic acid. Early reports were of its successful use as a single agent (Newman et al. 1962). However in 1965 resistance was being reported in isolates from individuals who had never been exposed to the drug (Evans and Waterworth 1966). Over the following decades investment in drug development declined (Silver 2011), but

with MRSA increasing broad-spectrum agents were increasingly required, as demonstrated by Vancomycin. In the early 1980s the quinolone ciprofloxacin was trialled in England (Crump et al. 1983). Early reports of studies from at this time in America demonstrated that the use of ciprofloxacin for eradication of MRSA colonisation was not effective and led to resistance (Mulligan et al. 1987; Peterson et al. 1990).

1.3.2 Beta-lactamase resistance in *S. aureus*

The introduction of Penicillin at the beginning of the 1940s heralded a new era in medicine. Systemic infection, which had a mortality of approximately 80% (Skinner and Keefer 1941), became treatable. Its introduction in 1940, was followed by reports of resistance as early as 1942 (Lowy 2003). Resistance had become pandemic by the mid-1950s (Rountree 1955), and present in the community by the 1960s (Lowy 2003). Before the drugs introduction into medical practice, *in vitro* analyses had shown that the drug could be inactivated by penicillinase enzyme encoded by *blaZ* (Abraham and Chain 1940; Kirby 1944). This gene is carried on either on a plasmid or a transposon. The beta-lactamase it encodes is synthesised upon exposure to beta-lactam antibiotics and subsequently hydrolyses the beta-lactam ring of penicillin leaving it inactive (Chambers and DeLeo 2009b). The development of resistance to what had been a miraculous drug led to the development and marketing of the semi-synthetic beta-lactam methicillin (Celbenin), that was insensitive to *blaZ* mediated inactivation, in 1959 (Knox 1960).

Methicillin, also known as Celbenin, was first used in 1959 in the UK (Knox 1960). Resistance to the new drug was identified less than a year later by a screening programme being undertaken at Public Health England. The newly observed resistance was identified because of pre-emptive assessment for resistance after the first usage of the drug, based upon the observations of how rapidly Penicillin resistance emerged (Jevons 1961). Of more than 5000

isolates screened during this study, only 3 were found to be resistant, but notably in individuals who had not received the therapy. These isolates shared common phage pattern (Group III), and carried a distinct multi-drug resistance profile (Methicillin/Penicillin/ Streptomycin/ Tetracycline) (Jevons 1961). Within 3 years invasive methicillin resistant *S. aureus* was isolated in Denmark (Eriksen and Erichsen 1964). These reports cumulatively represented the very first MRSA isolates ever identified, and what would later be identified as the first MRSA lineage (Crisóstomo et al. 2001; Enright et al. 2002).

1.3.2.1 Genetic basis of methicillin resistance: Staphylococcal cassette chromosome

The basis of methicillin resistance was not identified for more than 20 years after its first clinical detection. It is conferred by carriage of the *mecA* gene, encoding the penicillin binding protein 2a (PBP2a), a transpeptidase catalysing cross-linking of the peptidoglycan component of the bacterial cell wall (Hartman and Tomasz 1984; Matthews and Tomasz 1990). Carriage of *mecA* provides the recipient with an auxiliary PBP in addition to four already within the core genome. Expression of this PBP2a renders the recipient able to resist beta-lactams because of their low binding affinity to PBP2a (Hiramatsu et al. 2001). The recognition of the wide distribution of the *mecA* gene amongst staphylococci led to the consideration that it was being carried on a mobile genetic element (MGE) just as other resistance and virulence factors were. In 2000 that the mobile cassette of genes known as the SCC element was identified as the novel MGE carrying *mecA*, giving rise to MRSA (Katayama et al. 2000). Shortly after the initial description, variants of the element from differing MRSA strains were reported revealing their shared characteristics (Ito et al. 2001; Katayama et al. 2001).

The following summarises features, common to all of these SCC*mec* variants: the *mecA* gene within a mec gene complex, cassette chromosome recombinase (*ccr*) genes in a *ccr* gene complex, specific integration sites within the chromosome, which are termed integration site

sequences and, finally, repeat sequences flanking the integration site sequences (ISS) (Hiramatsu et al. 2001; IWG-SCC 2009) (Figure 1-3). These components and their variability between SCC*mec* variants now form the basis of the classification of the MGE (Ito et al. 2001). The following are used to differentiate the classes of element, and their subtypes.

(i) The *mec* gene complex

This is composed of *mecA*, the mec regulatory genes and flanking insertion sequences. The *mecA* gene is responsible for β- lactam resistance. The *mecI* and *mecR1* genes (Figure 1-3), upstream of *mecA*, encoding a transcription repressor protein and signal transduction protein respectively (IWG-SCC 2009). Together these are responsible for regulation of transcription of *mecA* in response to beta-lactam exposure. Downstream of these genes are an IS element and what is known as the hypervariable region (HVR). The *mec* gene complexes, which are used as part of the SCC*mec* classification system, vary on the basis of their *mecR1* genes, which IS elements are present and their orientation (IWG-SCC 2009).

(ii) The *ccr* gene complex

The chromosomal site-specific recombinase genes *ccrA* and *B, or C* with their surrounding open reading frames (ORF) constitute the *ccr* gene complex (Figure 1-3). These genes mediate the integration of SCC*mec* into the chromosome ensuring precise orientation and also any subsequent excision (Katayama et al. 2000). The *ccr* genes are defined on the basis of their DNA sequence similarity, being classified as A or B types for instance when there is between 50-85% nucleotide identity with other ccr genes in that category (IWG-SCC 2009).

(iii) Joining (J) regions

These are regions within the SCC*mec*, which lie outwith the *mec* and *ccr* gene complexes. Within the J regions other resistance determinants have been identified sub-classifying the major types further. The presence of genes, pseudogenes, transposons, or plasmids have been found amongst the varying types of element (IWG-SCC 2009).

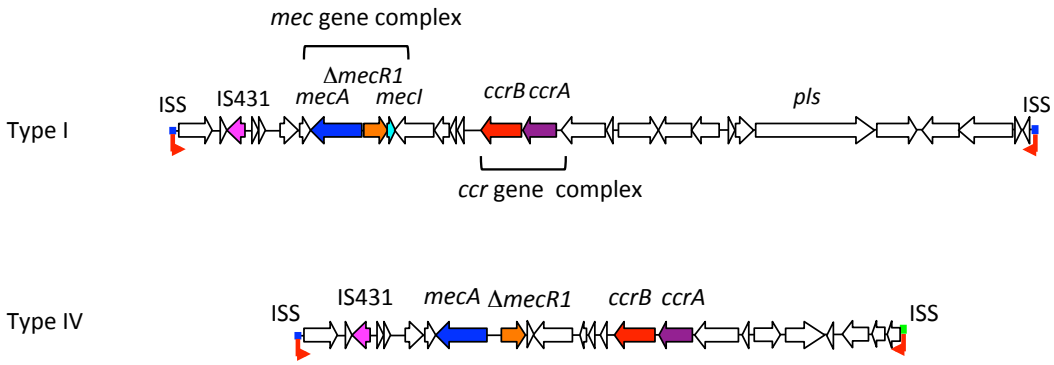

Figure 1-3 SCC*mec* components. Diagrammatic representation of main components of Type I element representative of the type found in *S. aureus* COL (Gill et al. 2005), and Type IV SCC*mec* found in MW2 (Baba et al. 2002). Coding sequences are coloured arrows as follows: blue (*mecA*), aqua (*mecI*), orange (*mecR1*), pink (IS*341*), white (other). Arrows indicate direction of transcription. ISS- integration site sequence. *pls*- methicillin resistant surface protein. Adapted from (García-Álvarez et al. 2011)

1.3.3 The emergence of the first MRSA and beyond

Following the descriptions of the SCC*mec* came the genetic studies characterising the strain background of the first MRSA clone. Using MLST typing of representatives of the first MRSAs identified in both the UK and Denmark it was shown that these were ST250, which belongs to *S. aureus* CC8 (Crisóstomo et al. 2001; Enright et al. 2002). This lineage carried a type I SCC*mec* element, as characterised by the SCC*mec* from NCTC 10442 (Ito et al. 2001). The type I element in this lineage is approximately 39kb in size (Figure 1-3), and had *mecA* as its only

resistance determinant unlike the type II and III SCC*mec* elements characterised at this point. It otherwise is notable for its carriage of the LPXTG surface protein *pls* (Ito et al. 2001). This large surface protein (230kDa) has been shown to function in bacterial aggregation and in binding to desquamated human epithelial cells (Huesca et al. 2002; Roche 2003). From its first emergence the ST250-MRSA-I lineage spread across Europe during the 1960 and 70s, but during the 1980s it became increasingly infrequent (Enright et al. 2002; Gomes et al. 2006). The single locus variant of the ST250-MRSA-I, ST247-MRSA-I, subsequently emerged and became successful both in Europe and America until the late 1990s (Roberts et al. 1998) when it too began to diminish, and both are now rarely reported (Monecke et al. 2011).

The gradual declining success of this first MRSA clone saw the emergence of newer and more successful lineages over the following decades. In the 1980s several new clones arose and became dominant primarily within the hospital setting. In the USA and Japan this was CC5 MRSA carrying a type II SCC*mec* element (Chambers and DeLeo 2009a). Across Europe and Australia during the same period, CC8 derived ST239 clone with a type III SCCmec, spread rapidly within hospital settings (Conceição et al. 2007; Willems et al. 2011). These respective SCC*mec* elements differed both in their respective resistance gene content, and size. For instance the Tn*554* present in the Type II SCC*mec* (Hiramatsu et al. 2001), or the pT181 and pUB110 in the Type III SCC*mec* element conferring bleomycin and tetracycline resistance. This multi-resistance phenotype is thought to one of the reasons behind their success in the nosocomial environment.

Following on from this in the early 1990s the CC30 lineage carrying a type II SCC*mec*, EMRSA-16 (ST36), became highly prevalent in the UK. By the year 2000 it accounted for up to 35% of the MRSA isolates identified through European Antimicrobial Surveillance System (Johnson et al. 2001). During this same period the CC22 EMRSA-15 (ST22) began to increase in prevalence, steadily over the 1990s into 2000s eventually overtaking the ST36 lineage which was declining

(Wyllie et al. 2011). EMRSA-15, notable for its carriage of a type IV SCC*mec* element, was initially reported in 1991 (Richardson and Reith 1993), but rose to prominence through the rapidity of its spread across UK hospitals, with very similar strains subsequently being reported globally (Ellington et al. 2010).

The type IV SCC*mec* has become a hallmark of the most successful MRSAs over the last two decades. Its success can also be measured in its sheer distribution across the clonal complexes. The type IV SCC*mec* has been found in 9 clonal complexes or STs (Chambers and DeLeo 2009b) in direct contrast to the limited ranges that its predecessors had entered. The type I, II and III SCC*mec* elements having been found in between 2 and 3 CCs respectively (Lina et al. 2006). Type IV elements are present both within hospital-associated lineages such as the ST22, but also with community associated clones such as USA300 that have become increasingly problematic since the late 1990s (DeLeo et al. 2010). The success of these lineages, HA and CA-alike, is in part thought to be due to the smaller SCC*mec* which they carry (Figure 1-3). Unlike their predecessors, these *mec* elements carry fewer resistance determinants, often only *mecA*. Type IV element carrying strains also appear to have a fitness benefit, with evidence showing that they have growth rates equivocal to MSSA strains (Okuma et al. 2002).

*1.3.3.1 The origins of SCC*mec

The exact origins of SCC*mec* remain unknown, but are strongly suspected to have derived from coagulase negative Staphylococcal species (CoNS). They are known to be widely distributed in staphylococci, including *Staphylococcus haemolyticus* and *Staphylococcus epidermidis* (Katayama et al. 2001). Their origin from CoNS has been hypothesised on the basis of several lines of evidence. Firstly that *mecA* homologues and methicillin resistance are extensively observed in CoNS, including *Staphylococcus sciuri* (Couto et al. 1996), *Staphylococcus capitis* (Schnellmann et al. 2006) and *S. epidermidis* (Dickinson and Archer 2000). The close identity of

the homologues of *mecA* in CoNS to those found in *S. aureus* suggests that there was a common ancestor amongst Staphylococci, which has been proposed to be *S. sciuri* (Couto et al. 1996). A single entry event of *mecA* into *S. aureus* was proposed by (Kreiswirth et al. 1993), with its subsequent propagation throughout the species. The entire element is hypothesised to have been amalgamated within CoNS backgrounds, then transferring into *S. aureus* (Otto 2013). In support of this are the high levels of sequence identity between SCC*mec* in *S. epidermidis* and *S. aureus* (Barbier et al. 2010). In addition to this, the fact that a type IV SCC*mec* element was found in *S. epidermidis* dating back to the 1970s (Wisplinghoff et al. 2003) decades prior to its first report in *S. aureus*, would fit with the element having originated in CoNS background. Whether this evolutionary process occurred entirely within *S. epidermidis* before its transfer to *S. aureus*, or if other components were derived from other CoNS however remains to be elucidated.

1.4 Atopic eczema and its association with *S. aureus*

As a pathogen, *S. aureus* is capable of causing a broad spectrum of disease both phenotypically and with respects to severity. Whilst we primarily study this organism for its infectious potential, it can also contribute to disease through inducing inflammatory responses in the host. The highly prevalent skin disease atopic eczema is the clearest example of this, where the interaction of the organism with the skin in affected individuals contributes to cutaneous inflammation.

1.4.1 Atopic eczema: Prevalence and pathogenesis

Atopic eczema (AE) is the commonest inflammatory skin disease of childhood, affecting up to 25% of children in the UK (Shamssain 2007). The onset of AE often precedes other atopic diseases including food allergy and asthma, with successive acquisition of these related conditions known as the atopic march (Zheng et al. 2011). It is a complex disease characterised

by epidermal barrier dysfunction, inflammation of the skin and propensity to develop allergic Th2 type responses (Irvine et al. 2011). The underlying genetic predisposition to develop this disease is heavily influenced by immunological and environmental factors such as allergen exposure, diet and microbes (Roduit et al. 2012; Leung and Guttman-Yassky 2014). Individuals with AE are strikingly prone to colonisation by *S. aureus*, rates of colonisation on eczema affected skin being 90% in comparison to 5% of normal skin in healthy individuals (Leung and Guttman-Yassky 2014).

Normal skin functions as a series of interconnected barriers that work to retain moisture, and to prevent infiltration by environmental allergens and pathogens. The epidermis constitutes the outermost layer of the skin and is the first line of defence against these potentially harmful influences (Figure 1-4). It is formed through progressive and organised terminal differentiation of keratinocytes starting from the basal layer (*stratum basale*) extending gradually to the outermost layer, the *stratum corneum* (SC). The outer layer is a compact, tough multi-layer of anucleate cells embedded in a lipid rich extracellular matrix. The lipid content, hydrophobicity, and organisation into a series of lamellar bilayers, make the SC a strong barrier preventing moisture loss and invasion of microbes (Elias 2005). Several of the constituent lipids of this layer also exhibit antibacterial actions *in vitro* including free fatty acids and ceramides (Miller et al. 1988). Below the SC is the *stratum granulosum* in which cells are cohesively bound by intracellular tight junctions. Together the SC and SG form a physical barrier. (McAleer and Irvine 2013)

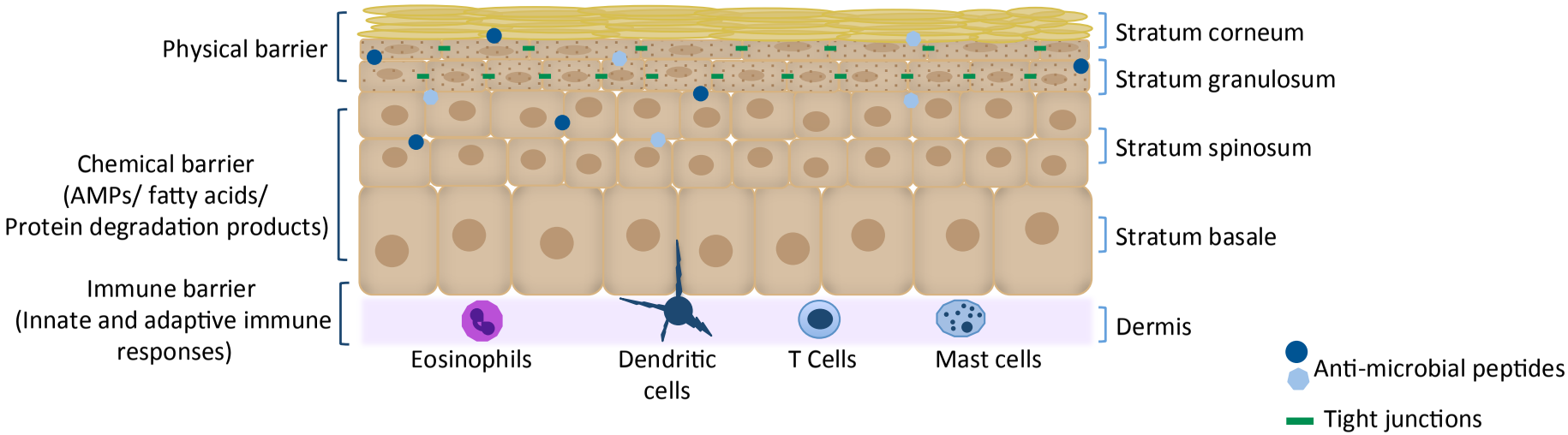

Figure 1-4 The human skin barrier. The layers of the epidermis (labelled on right) progressively stratify from the basal layer (stratum basale) upwards to form the cornified layer (stratum corneum). Together the stratum corneum and stratum granulosum layer below form the physical barrier, which protects the skin from moisture loss and microbial invasion. These layers along with the stratum spinosum form a chemical barrier, where antimicrobial peptides, and epidermal protein breakdown products (such as filaggrin) inhibit microbial growth. Below the stratum basale is the dermis, where the adaptive and innate immune cells are resident. (Figure is adapted from Kuo et al. 2013). (AMPs- antimicrobial peptides). Different shades of blue are indicative of differing subtypes of AMPs such as Human beta defensins and cathelicidin (colour is not specific to type).

In addition to the physical barrier, these layers also provide protection as a chemical barrier. The keratinocytes of the epidermis produce antimicrobial proteins including the antimicrobial peptides (AMPs) human beta defensins and cathelicidin.The breakdown products of constituent proteins of the epidermal cells are additionally used to moisturise and control the pH of the epidermis (Kuo et al. 2013). If this barrier becomes breached then allergens and pathogens interact with the immune barrier of the skin, the innate immune system. Keratinocytes of the skin display an array of pattern recognition receptors such as Toll like receptors (TLR). Stimulation of these receptors leads to production of antimicrobial peptides, which are both antimicrobial and stimulate strengthening of the tight junctions to prevent further penetration by microbes or allergens (Yuki et al. 2011). Failing this stimulation of immune cells within the dermal layer will occur triggering both innate and adaptive immune responses.

In AE defects in each of the aforementioned protective mechanisms have been shown to be contributory to the disease pathogenesis. The effectiveness of the barrier is breached by deficiencies of structural proteins of the SC such as involucrin, loricrin and filaggrin (Palmer et al. 2006), as well as reduced lipid content (Leung and Guttman-Yassky 2014). Tight junction defects have also been shown in AE patients, further contributing to the deficient epidermal barrier (Elias and Wakefield 2014). The skin of individuals with AE produces less AMPs in comparison to normal individuals as well as those affected by psoriasis, another common inflammatory skin disease with abnormal epidermal barrier (Ong et al. 2002). The underlying immune disorder in AE is complex and known to negatively influence the functioning of the epidermal barrier. During acute inflammation in AE skin there is activation of the Th2 and 22 immune pathways, leading to production of cytokines such as IL-4, 13, and 31 (Gittler et al. 2012). These in turn have been shown to reduce AMP expression and expression of genes important for terminal epidermal differentiation and structure of the epidermis including filaggrin (Boniface et al. 2005). Several of these factors have been suggested to contribute to

the enhanced colonisation by *S. aureus* seen in AE, these will be considered individually in the following section.

1.4.2 Enhanced colonisation by *S. aureus* in AE

The factors underlying the increased rates of *S. aureus* carriage are thought to be multifactorial. Interestingly nasal colonisation rates in AE are not as high as cutaneous carriage rates, but these are variably reported across the studies assessing this (Totté et al. 2016). A recent study showed that neither increased transepidermal water loss, a marker of epidermal barrier dysfunction, or AE associated with filaggrin mutations are associated with increased nasal or faucal carriage of *S. aureus* (Berents et al. 2015). This suggests that the enhanced carriage rates are due to the disease process occurring at non-nasal sites. Several features of this disease are thought to contribute to the observed increase in colonisation by *S. aureus*. For instance the reduced lipid content, overall dryness and in some instances cracked nature of the skin in AE is thought to facilitate colonisation. The following are the disease-associated features proposed from previous studies to contribute to this process.

1.4.2.1 Enhanced binding

In vitro comparison has revealed enhanced binding of *S. aureus* to the skin of individuals with atopic eczema, in comparison to those with psoriasis or normal skin (Cho et al. 2001a). The authors of this study demonstrated that *S. aureus* was adhering to the SC from human skin sections from AE, normal and psoriasis-affected individuals. Immuno-histochemical analysis revealed that in AE skin, including unaffected sites, this due to the redistribution of fibronectin to the SC in comparison to the other patient groups (Cho et al. 2001a). From this they demonstrated that *S. aureus* mutant strains, missing cell wall anchored proteins (CWA) including clumping factors A and B and fibronectin binding proteins A and B, bound

significantly less to AE skin in comparison to psoriatic or normal skin (Cho et al. 2001a). Conversely they demonstrated that the presence or absence of collagen adhesin (*cna*) in *S. aureus* strains had no impact on binding of the organism to AE skin. These findings suggested that increased expression of fibrinogen and fibronectin, as tissue ligands of the CWAs, in the context of AE associated inflammation contributed to the enhanced colonisation. Finally this evidence supported observations reported in murine models of skin inflammation, where *S. aureus* binding to the skin of mice with Th2 type skin inflammation (such as would be characteristic of AE) was significantly increased (Cho et al. 2001b).

1.4.2.2 Impaired innate immune barrier

The immune barrier of atopic skin has been shown to be impaired at several levels. In AE there is reduced production of the antimicrobial peptides which is believed to contribute to the increased risk of overgrowth of *S. aureus*, and also the increased risk of development of herpes simplex infection in a sub-group of AE patients (Ong et al. 2002). Expression of human β defensin 2, which has activity again gram-positive bacteria, and the cathelicidin LL-37 which has both antibacterial and antiviral activity, is reduced in atopic skin compared with normal and psoriatic skin (Ong et al. 2002). In addition to reduced expression, the cytokines produced by Th2 cells in AE, including IL(interleukin)-4 and 13, have been shown to inhibit the production of AMPs as well as TLR-stimulated production of these peptides (Howell et al. 2006; Kuo et al. 2013). The extent to which altered AMP production contributes to the increased susceptibility to colonisation and infection in AE is still not fully understood. For instance HBD (human beta defensin) 1 and 2 have been reported to have no activity *in vitro* at physiological salt concentrations, making their functionality in the context of AE much less clear (Otto 2010).

1.4.2.3 Altered pH

In healthy skin the pH is reported is reported to range between 4 and 6 (Rippke et al. 2004; Kim et al. 2016). This is maintained through sweat, presence of free fatty acids, and breakdown products of both phospholipids and filaggrin. However, all of these are reduced in AE with subsequent increase in pH of the skin thought to favour increased *S. aureus* growth (Rippke et al. 2004).

The by-products of filaggrin degradation, urocanic and pyrrolidone carboxyilic acid (UCA/ PCA), are key to providing moisturisation of the normal skin and control of its pH (Brown and McLean 2012). At physiological concentrations both of these products have been shown to reduce *S. aureus* growth rates in culture (Miajlovic et al. 2010). The authors of this study demonstrated that acidification of growth media to a pH of 5.5 led to reduced expression of key *S. aureus* surface proteins involved in colonisation including ClfB, FnBPA and Spa. Notably the expression of another surface protein involved in adherence to keratinocytes, IsdA, was shown to have reduced expression in the presence of UCA and PCA, which was not related to pH (Miajlovic et al. 2010). Whilst these findings were seemingly more relevant in the context of AE associated with underlying filaggrin mutations they demonstrated the potential importance of the altered skin pH towards alkalinity in AE for *S. aureus* colonisation.

1.4.3 How does *S. aureus* contribute to disease activity in AE?

Much of what is understood of the association between this organism and AE is derived from clinical observational evidence. Decades of study have reported high *S. aureus* carriage rates as well as high bacterial burden on both AE affected and unaffected skin (Leyden et al. 1974; Aly et al. 1977; Williams et al. 1990; Hoeger et al. 1992; Guzik et al. 2005). Disease severity in AE

has been shown to correlate to the bacterial burden at lesional sites, as well as the total number of body sites colonised (Lomholt et al. 2005; Tauber et al. 2016). This link with burden and severity cannot simply be explained as infection. Colony burdens of up to 10^7 CFU (colony forming units)/ cm^2 can be recovered from the skin of AE individuals without signs of overt infection such as crusting or weeping (Leyden et al. 1974). The organism contributes to the inflammation through a variety of proposed mechanisms, many of which are not fully understood and are discussed in the following sections.

The majority of the work that has been undertaken to understand how this organism contributes to AE disease severity has focussed on target products of the organism, and the specific immune responses that they elicit. The most extensively studied *S. aureus* virulence factors in this disease are toxins. The following summarises bacterial components that have been implicated in the induction of inflammation and disruption of the epidermal barrier.

1.4.3.1 Toxins

The staphylococcal enterotoxins (SE) A, B, C, D and TSST-1 have been shown to function as so-called superantigens in eczema, because of their ability to stimulate polyclonal T cell activation and cutaneous inflammation (Ong and Leung 2010; Spaulding et al. 2013). These toxins also have the ability to stimulate the release of histamine causing allergic type skin inflammation (Bunikowski et al. 2000; Ardern-Jones et al. 2007). Enterotoxin B has specifically been shown to induce production of the inflammatory cytokine IL-31 which has been suggested to contribute to chronic pruritus, a prominent feature of AE (Sonkoly et al. 2006). *In vitro* evidence has shown that enterotoxins are able to alter the expression of glucocorticoid receptors on T cells, inducing glucocorticoid insensitivity which is thought to be a mechanism contributing to the failure of topical steroid therapy in a subgroup of severe AE cases (Hauk et al. 2000; Schlievert et al. 2010). Clinical evidence suggests that detection of toxins and

superantigen specific IgE in AE patients are both associated with greater disease severity (Bunikowski et al. 2000; Zollner et al. 2000; Capoluongo et al. 2001). Detection of these toxins in AE isolates ranges in the literature between 50 and 80% (Bunikowski et al. 2000; Ong and Leung 2010), whilst SA specific IgEs ranges between 33-96% of studied AE cases (Bunikowski et al. 2000; Zollner et al. 2000). While functional evidence has shown these toxins are contributing to the inflammatory process in eczema, the clinical association is not as convincing given reported lack of strong correlation between toxin production or IgE presence with disease activity (Bunikowski et al. 2000; Arkwright et al. 2001). These enterotoxins are also variable in their presence on the basis of strain background, which has not been taken accounted for in previous studies. Additionally SEB has been shown to stimulate eczematous change in normal skin (Strange et al. 1996). This would suggest that they are contributory to disease activity rather than causal, as some studies claim.

Several notable studies of the haemolysins have provided functional insight in to the organism's contribution to the disease. For instance Haemolysin α is cytotoxic to keratinocytes and has been shown to stimulate cytokine production in AE lesions (Breuer et al. 2005; Brauweiler et al. 2014). Filaggrin deficiency in AE is believed to make the affected individuals more susceptible to the α-toxin induced cytotoxicity (Brauweiler et al. 2013). Strong functional evidence recently emerged showing that haemolysin δ directly stimulates mast cell degranulation, production of IgE and IL-4, as well as allergic type skin inflammation in a murine skin model of *S. aureus* colonisation (Nakamura et al. 2013). This finding represents the clearest evidence to date of direct triggering of AE disease to date. These candidates are arguably more likely contributory to the disease than the superantigens given that they are highly conserved across the species. However their expression can be both vary according to their clonal background, and are dependent upon functioning global virulence regulatory systems such as *agr* and *saeR/S* (Li et al. 2009b; Montgomery et al. 2010; Monecke et al. 2014a).

1.4.3.2 Cell surface proteins

Although surface proteins have been the focus of less research in AE several have been shown to induce cutaneous inflammation. Peptidoglycan from *S. aureus* has been shown to stimulate the release of thymic stromal lymphoprotein from keratinocytes, which is a potent inducer of allergic inflammation, and mast cell infiltration of the dermis in a murine model (Allakhverdi et al. 2007; Matsui et al. 2007). Similarly lipoteichoic acid has been shown to stimulate allergic type cytokine production in murine skin and human explant skin (Matsui and Nishikawa 2003; Travers et al. 2010). The surface proteins Spa (Surface Protein A) and FnbpA/B (Fibronectin binding proteins) have both been shown to induce allergic type inflammatory responses in murine models (Terada et al. 2006; Reginald et al. 2011). Moreover the fibronectin binding proteins have also been proposed to have superantigen like activity, with identification of IgE from patients with AE cross reacting with *FnbpA* (Reginald et al. 2011).

1.4.3.3 Secreted enzymes

S. aureus secretes a wide variety of enzymes which function to weaken or destroy host molecules as well as impeding other processes such as signalling cascades (Otto 2014). The proteases are an example of this and are used by the organism to disrupt host tissues and aid dispersal in the host. In AE, the serine V8 protease, *sspA*, has been implicated as a potential mechanism by which colonisation by *S. aureus* damages the skin. Specifically, application of the protease to the skin of nude mice caused disruption of the *stratum corneum* (Hirasawa et al. 2010).

1.4.4 Bacterial factors and clinical epidemiology

Clinical studies in this area have identified several other aspects of the carriage of this organism, which may be of importance in understanding the relationship between *S. aureus*

and disease activity in AE. Broader aspects of the organism in association with AE such as duration of carriage, source of colonisation, and strain types associated have been much less studied. They are demonstrative of areas needing further investigation. These epidemiological studies are summarised in the following sections.

1.4.4.1 Strain persistence and transmission

Nasal carriage studies frequently assess the temporal aspects of carriage, and over extended periods. This has however, has not been the case in AE research with respects to *S. aureus*. There are two exemplar studies which have undertaken longitudinal follow up of children with AE for a mean of 9 and 12 months respectively (Hoeger et al. 1992; Lomholt et al. 2005). Sampling of 20 children over a period of 9 months by Hoeger et al. revealed that in 70% of children strains of the same phage and capsule type were re-isolated (Hoeger et al. 1992). With sequential sampling of 11 children over the course of 1 year, Lomholt *et al* showed that 45% of these children were colonised on each of the first 8-monthly visits. The PFGE (pulsed field gel electrophoresis) typing method used by this study revealed one child carrying a strain of the same PFGE type during each of the 9 sampling episodes over 9 months, with four other children carrying strains of the same PFGE type over 4 successive sampling episodes (Lomholt et al. 2005). Interestingly, the single case carrying the same PFGE type for 9 months (child 2) had the overall greatest disease severity, and severe disease was noted at each of the 9 visits (Lomholt et al. 2005). The authors of both studies suggest that prolonged carriage may be contributory to disease, but no association was made with severity.

The primary source of colonisation in AE cases has not generally been addressed by previous studies. Spread of the same strain from AE affected skin to unaffected skin has been shown in two prior studies, with colonisation at non-inflamed skin sites thought to be a later event relating to the relative increase in bacterial burden with disease flare (Hoeger et al. 1992;

Lomholt et al. 2005). Colonisation simultaneously at nasal and extra-nasal sites by the same strains on basis of PFGE, phage type or resistance toxin profile has been shown (Hoeger et al. 1992; Capoluongo et al. 2001; Pascolini et al. 2011). However, none of these studies have concluded as to which event arose first, nasal or extra-nasal colonisation. Inter-family transmission has also been shown, with one study showing that 38% of AE patients transmitted *S. aureus* to their mothers (Hoeger et al. 1992). In a separate study, concomitant carriage of the same strain in mother and children with AE was shown, but with unknown directionality of spread (Pascolini et al. 2011). Transmission within the family has been postulated to be one of the reasons underlying re-colonisation immediately after cessation of eradication therapy in AE (Lever et al. 1988).

1.4.4.2 Strain prevalence or change during colonisation

Typing of *S. aureus* strains in AE studies have used a number of methodologies making subsequent comparison between these works difficult. Several authors have reported a lack of a prevailing strain background in association with carriage in AE or disease severity (Kim et al. 2009; Yeung et al. 2011; Kong et al. 2012). Interestingly four previous studies found that CC30 isolates were identified at lower frequency in AE cases than in nasal carriage controls (Kim et al. 2009; Balma-Mena et al. 2011; Yeung et al. 2011; Rojo et al. 2014).

Whilst strain type has not been associated with colonisation in AE, change of strain during colonisation has been reported as being significantly associated with increased disease severity (Lomholt et al. 2005). During longitudinal follow up children who became colonised by a new strain of differing *agr* type had a significant flare of their disease. In 3 out of 4 instances where this was observed, the affected child's disease had been improving prior to the presumed novel colonisation episode (Lomholt et al. 2005). The associated deterioration in disease control was postulated to have been due prior down-regulation of virulence in the

previous colonising strain or production of new virulence determinants by the new strain. No other examples of a similar nature have been reported in the literature.

1.4.4.3 Antimicrobial resistance

Reported rates of methicillin resistance are highly variable in individuals with AE and influenced by the country of origin of the study. They range from 0.5% in studies from North America to 18% in South East Asia and India (Chung et al. 2008; Lo et al. 2010; Balma-Mena et al. 2011; Jagadeesan et al. 2014). Colonisation by MRSA at AE sites has not been clearly demonstrated as associated with disease, but has been shown to increase the risk of subsequent development of skin and soft tissue infection (Lo et al. 2010). Increased number of superantigens produced by MRSA isolates compared with MSSA isolates in isolates has been proposed as a potential mechanism by which AE disease severity might be increased with MRSA colonisation, but this is still speculative (Schlievert et al. 2010). Finally resistance to other antimicrobials have not been specifically linked to disease activity or severity.

1.4.4.4 Presence or absence of genes

Association between AE disease and the presence of specific genes in *S. aureus* isolates is more commonly investigated in this field. Several of these target genes have been reported as being significantly associated with disease severity. These include the presence of *SEB* (Capoluongo et al. 2001), *PVL* (Panton-Valentine Leukocidin) (Lo et al. 2010), *TSST* (Toxic Shock toxin)(Pascolini et al. 2011), *sasG* (S. aureus surface protein G) and *SCN* (Staphylococcal complement inhibitor) (Rojo et al. 2014). Multiple other genes have, however, been reported as being more commonly present within AE disease-associated isolates including *ssl8* (superantigen like protein 8), *splB* (serine protease B) *lukD*, *lukE* (Rojo et al. 2014), *SEA* (Na et al. 2012), *SEB*, and the *egc* (enterotoxin gene) cluster (Schlievert et al. 2008). These genes have

largely not been interrogated with respects to functional contribution to AE activity with the exception of the superantigens.

1.5 Other microbes in AE

Progressively the focus in AE research is shifting towards how other microorganisms may be involved in the aetiology of the disease. There is a well-known association with increased susceptibility to cutaneous viral infection in AE. The classical example of this being Eczema Herpeticum (EH), which is widespread infection of atopic skin with herpes simplex virus (HSV) 1 or 2. This is considered a severe complication of the disease because of the potential for it to cause systemic disease, including pneumonia and encephalitis (Dreyfus 2013). However, unlike the association with *S. aureus* in eczema, this is significantly less common. Approximately 3% of AE patients are predisposed to this complication, with those at risk factors including early onset severe disease, high systemic levels of IgE levels and co-morbid atopic diseases such as food allergy (Beck et al. 2009; Dreyfus 2013). It is also increasingly thought to represent a sub-group of patients, with a different disease endotype relating to the underlying genetic and immune basis of their disease (Leung and Guttman-Yassky 2014).

The first descriptions of the cutaneous microbiome revealed specific variances of microbial population structure according to the skin site (Grice et al. 2009) and type of skin at those respective sites, such as sebaceous, moist or dry skin. The predilection for AE to affect very specific body sites such as the antecubital and popliteal fossae led the investigators to assess changes in the microbiome at these sites in children with AE throughout a cycle of disease quiescence, flare and recovery (Kong et al. 2012). The baseline comparison between disease predilection sites in AE cases and healthy controls revealed widely differing microbial populations between the groups suggesting that cutaneous microbiome was perturbed in AE disease even without active disease. During disease flares these authors demonstrated that

overall cutaneous microbial diversity in AE cases was markedly reduced, with overall expansion of staphylococcal species. The predominant increase was accounted for by *S. aureus*, which reaffirms the association with disease flares and the organism (Kong et al. 2012). There was also, however, a significant increase in *S. epidermidis* during flares that was thought to potentially reflect a compensatory mechanism to control the *S. aureus* population. Following on from this work it has recently been shown that commensal staphylococcal colonisation seemingly has a protective effect in infancy against the development of AE in the first year of life (Kennedy et al. 2016).

1.6 Overall aims of the study

The overarching aim of this project was to understand the genetic basis of this organism's (*S. aureus*) adaptation in the face of human intervention and influence on survival in the human host, specifically during colonisation of the skin. The aims of each chapter presented in this thesis were as follows:

Chapter 3: Genomic Investigation of the emergence of the first Methicillin resistant *Staphylococcus aureus* (MRSA).

In 1960 the first incidence of MRSA was identified only one year after the introduction of the drug. Using WGS of a collection representative of the very first MRSA isolates identified in Europe between the 1960s and 1980s this work aimed to trace the evolutionary events leading to the emergence of the very first MRSA lineage, and to estimate the historical time point at which the SCC*mec* was first acquired by *S. aureus*.

Chapter 4: *Staphylococcus aureus* carriage in healthy children

This work aimed to characterise the genetic diversity of strains colonising healthy children in the community, looking for evidence of adaptive changes to the host. This analysis served as a control population for comparison with AE cases assessed in Chapter 5.

Chapter 5: *Staphylococcus aureus* associated with Atopic Eczema prospective case study

The aim of this study was to assess the genetic heterogeneity of the colonising strains of children with active moderate to severe eczema, looking for evidence of micro-evolutionary adaptive changes arising in the host, which may be of biological relevance in the disease.

Chapter 6: *Staphylococcus aureus* in Atopic Eczema: case control study

Comparative genomic analysis was undertaken on a collection of isolates from healthy nasal carriers and AE cases with the aim of determining the population structure of the collection, as well looking for evidence of variable gene content which may be associated with disease.

2 Materials and methods

2.1 Study design

2.1.1 Ethical and local authority approval

Results presented in Chapters 4, and 5 are derived from clinical studies involving human participants. Prior to commencement these studies were reviewed by the institutional sponsor on behalf of the University of Dundee as well as committees of the National Research Ethics Service. Approvals were granted by the Nottingham 1- East Midlands Research Ethics Committee (14/EM/1299- *Staphylococcus aureus* associated with Atopic eczema; Chapter (4)) and the East of Scotland Research Ethics service (15/ES/0153- *Staphylococcus aureus* carriage in Tayside children; Chapter (5)) respectively.

Community control sampling was undertaken at schools and nurseries within the Tayside and North Fife area. Written permission for study participation at these sites was obtained from the relevant local authorities and/ or nursery management following their review of the study.

Case Samples pertaining to *S. aureus* associated with atopic dermatitis flares: case control study in Chapter 6 were collected by Professor Alan Irvine and Dr Maeve McAleer who kindly gifted this sample collection for the work contained in this thesis. The samples were obtained following approval of the study by the Research Ethics Committee of Our Lady's Children's Hospital, Crumlin, Dublin, Ireland. Control samples for this study were collected and graciously provided by Dr Désirée Bennett. Ethical permission for collection of these samples was provided by the Temple Street University Hospital Ethics Committee, Dublin, Ireland.

All studies were conducted in accordance with the principles of the declaration of Helsinki. Prior to participation written parental consent was obtained as well as child assent. Ethics and local authority approval letters for these studies are accessible online via: https://figshare.com/ (Login details are as per Page x- supplementary material).

2.1.2 Recruitment criteria

The following descriptions pertain to sample collection studies presented in **Chapters 4 and 5** only. The sample collection ultilised for analysis in Chapter 6 was provided by our collaborators Professor Alan Irvine and Dr Desiree Bennett as stated above (Section 2.1.1). Details of the study criteria used to obtain this sample collection are presented in Chapter 6.

Atopic eczema case study participants were recruited on the basis of active skin disease on the day of participation. This was determined by clinical history and examination by Dr Catriona Harkins. Disease severity was scored according to the Eczema Area Severity Index (EASI) score (Hanifin et al. 2001) in accordance with current recommendations of standardised reporting in AE clinical studies (Schmitt et al. 2014).

Case study inclusion criteria were: age 0-8 years, moderate to severe AE, and written parental consent. Exclusionary criteria were: antibiotic (oral or topical) less than 4 weeks previously, topical antiseptics therapy within 2 weeks, and ultraviolet therapy (UV) of any type within 3 months.

Control study participant inclusion criteria were: age 0-12 years, written parental consent, and child assent. No exclusionary criteria were applied to participation in this study, but samples obtained from participants meeting the exclusionary criteria of the case study were not used as comparators to AE cases. Explicitly this meant controls selected were: below 8 years of age, with no history of AE, or recent antimicrobial therapy (as defined for case selection).

2.2 Media and reagents

2.2.1 Media and bacterial growth conditions

All *S. aureus* strains were routinely grown in Brain Heart Infusion (BHI) media (beef heart infusion solids 5 g/L, brain infusion solids 12.5 g/L, disodium phosphate 2.5 g/L, glucose 2 g/L, proteose peptone 10 g/L, disodium phosphate 2.5 g/L) (Oxoid, UK) or on BHI supplemented with 1.5% bacteriological agar (Oxoid, UK). For confirmation of haemolysis activity colonies were grown on pre-purchased Columbia agar with 5% sheep's blood (Oxoid, UK).

Strain stocks were derived from single colony isolates. Briefly, single colonies were used to inoculate 5 ml BHI media and grown overnight at 37°C whilst shaking at 200 rpm. 700 µl of overnight culture was then added to an equal volume of heat sterilised 40% glycerol and stored at -80°C. All subsequent recovery was by subculture from -80°C colony stock into 5 ml BHI with overnight growth at 37°C whilst shaking at 200 rpm.

2.2.2 Reagents

All chemicals used for this work were of analytical grade. The buffers and solutions used are listed in Table 2-1. Other buffers and solutions used for this study were derived from pre-purchased kits and are detailed in the following relevant method sections.

Table 2-1 Reagents used for this study.

Buffer/ Solution	Components
Tris-acetate EDTA (TAE) (Thermo Fisher Scientific, UK)	40mM Tris acetate, pH 8.3 1mM EDTA
Phosphate buffered saline (PBS) (Gibco Life Technologies, UK)	10mM $NaHPO_4$ 2.68mM KCl 0.14M NaCl
Tris HCl (Sigma Aldrich, UK)	1.0 M Tris-HCl pH 7.4
DNA loading dye (Thermo Fisher Scientific, UK)	10mM Tris-HCl 0.03% (w/v) Bromophenol blue 0.003% (w/v) Xylene cynanol FF 60% (v/v) Glycerol 60mM EDTA

2.3 Isolation and identification of S. aureus *from clinical samples*

2.3.1 Swabs and sampling

Skin swabs (Transtube Amies swab, Medical Wire, England) were obtained by rubbing the swab in a circular fashion on sampling site for 10 seconds. Cases were sampled across multiple body sites including two eczema-affected areas, two clinically unaffected areas and a single nostril. Controls were swabbed from a single nostril and a single antecubital fossa. Swabs were stored at 4°C prior to processing within a 48 hour period to prevent commensal overgrowth.

2.3.2 Isolation of *S. aureus* from clinical swabs

**Components of the following method have been adapted from the unpublished protocol courtesy of Dr Ewan Harrison, Department of Veterinary Medicine, University of Cambridge.*

Two alternative methodologies (pre- and post- enrichment) were tested and compared for the cultivation of bacteria from uncharacterised clinical isolates.

Swab tips were removed using flame-sterilised scissors and placed in 500 µl 1x PBS, and incubated at room temperature for 30 minutes. For non-enriched cultivation 100 µl was plated on Brilliance Staph. 24 chromogenic agar (Oxoid, UK), and incubated in air at 37°C for 24 hours. For enrichment 100 µl of the swab fluid was used to inoculate 3 ml of Nutrient Broth media with 7.5% NaCl (Oxoid, UK), incubated statically in air at 37°C for 18 hours. A 100 µl aliquot of this broth was then plated on chromogenic agar and incubated as above.

After 24hours of incubation plates were assessed for growth. A colony score was given to each plate on recovery of blue colonies (presumed *S. aureus*) as follows: 0= no growth, 1= 1-20, 2= 20 – 100, 3= >100. Negative plates were re-incubated for a further 24 hours and reported as

negative if no growth was observed at this point. Presumptive *S. aureus* colonies were then sub-cultured onto BHI agar, and incubated in air at 37°C for 16 hours. A maximum of 10 colonies were sub-cultured onto BHI per swab sample per body site.

2.3.3 Colony PCR

Colonies were confirmed as *S. aureus* by PCR detection of the species specific *femB* and *nuc* genes as previously described (Zhang et al. 2004; Paterson et al. 2012). Specifically single colonies were lifted from BHI subculture plates and re-suspended in 50 µl of heat-sterilised MiliQ water by mixing. Cells were then lysed by heating to 95°C for 15 minutes whilst shaking at 1,400 rpm on a micro-tube mixer (Thermo-mixer, Eppendorf). Cell debris was then removed by centrifugation at 10,000 rpm (9391 x G) for 3 minutes (Eppendorf 5424 Microcentrifuge; Eppendorf, Germany). A 2 µl aliquot was then used as the DNA template for the subsequent PCR reaction.

DNA was amplified using TopTaq DNA polymerase (Qiagen, UK) according to the manufacturers protocol. Primers used for this study were synthesised by (Eurofins Genomics, UK) with sequences listed in Table 2-2. Reactions were prepared to a final volume of 25 µl and as follows: 12.5 µl TopTaq master mix (containing final concentration of 1.25 u TopTaq DNA polymerase, 1x TopTaq PCR buffer (Qiagen, UK), 200 µM of each dNTP, 1.5mM $MgCl_2$; Qiagen), 2 µl of primers (final concentration of 0.4 µM), 2.5 µl TopTaq coral load gel loading reagent (10 x concentrate), and 6 µl RNAase free water (Qiagen, UK). The following PCR condition were used for all reactions: initial denaturation 94^{o}C for 3 minutes, 28 cycles of 94^{o}C for 30 seconds, 50^{o}C for 30 seconds, 72^{o}C for 1 minute, and final extension time of 10 minutes at 72^{o}C.

The DNA products were then resolved by gel electrophoresis using 1% agarose gels. In brief, 6 µl of PCR product was suspended in 1 µl in 6x DNA loading dye (Thermo Fisher Scientific, UK) and

loaded into wells in 1% agarose gel prepared with 1x Tris acetate-EDTA buffer with 1x SyberSafe DNA gel stain (Thermo Fisher Scientific, UK). Electrophoresis was performed in 1x TAE buffer under constant voltage of 120V for 40 minutes. Nucleic acids were then imaged using a ChemiDoc XRS+ system (BioRad, UK).

Table 2-2 Primers used in this study. Nucleotide positions are relative to *femB* (locus ID: SAR1388) and *nuc* (locus ID: SAR0847) in *S. aureus* reference genome MRSA252 (accession number: BX571856).

Primer	Sequence	Product length (bp)	Nucleotide Position	Reference
femB (F)	CATGGTTACGAGCATCATGG	533	415-434	(Paterson et al. 2012)
femB (R)	AACGCCAGAAGCAAGGTTTA		926-945	(Paterson et al. 2012)
nuc (F)	GCGATTGATGGTGATACGGTT	279	286-306	(Zhang et al. 2004)
nuc (R)	AGCCAAGCCTTGACGAACTAAAGC		541-564	(Zhang et al. 2004)

2.4 DNA methods

2.4.1 Extraction

Genomic DNA was extracted from overnight cultures of single colonies grown for 16 hours at 37°C using the Masterpure™ Gram Positive DNA purification kit (Epicentre, Illumina, UK). A 700 µl aliquot of the overnight culture was pelleted by centrifugation at 10,000 rpm for 10 minutes (9391 x G) (Eppendorf 5424 Microcentrifuge; Eppendorf, Germany). The supernatant was discarded by pipetting and cell pellet re-suspended in 150 µl of Tris-EDTA buffer, with 1 µl of Ready-Lyse lysozyme (Epicentre, Illumina, UK) and addition of lysostaphin (Sigma Aldrich, UK) to a final concentration of 50 µg/ml. Cell lysis proceeded by incubation at 37°C for 30 minutes. Following this proteinase K was added to a final concentration of 50 µg/ml along with 150 µl Gram-positive lysis solution (Epicentre, Illumina, UK), and vortexed vigorously for 10 seconds. Samples were then incubated at 65°C for 15 minutes, with vortexing at 5 minute intervals for 5 seconds. Samples were cooled to 37°C before being placed on ice for 5 minutes. Protein

precipitation followed by addition of 175 µl of MPC protein precipitation reagent (Epicentre, Illumina, UK) and vortexing for 15 seconds. Cell debris was pelleted by centrifugation at 4°C for 10 minutes at 13,000 rpm (15,871 x G) (Eppendorf 5424 Microcentrifuge). The supernatant was then transferred to a fresh microcentrifuge tube and treated with 1 µl RNase A (final concentration of 5 µg/µl) incubated at 37°C for 30 minutes. To the supernatant 500 µl of isopropanol was added before being inverted 30 times. DNA was pelleted by centrifugation at 4°C for 10 minutes at a speed of 13,000 rpm (15,871 x G) (Eppendorf 5424 Microcentrifuge). Isopropanol was then removed by pipette and DNA pellet washed with 500 µl of 70% ethanol. DNA was again pelleted by chilled centrifugation for 5 minutes at 13,000 rpm (15,871 x G) (Eppendorf 5424 Microcentrifuge). Following removal of ethanol DNA pellets were air dried at room temperature for 20 minutes and then re-suspended in 35 µl of TE buffer.

2.4.2 Quantification and quality assessment of DNA

Concentration of extracted gDNA was then quantified using Qubit dsDNA broad range assay kit (Life Technologies) as per the manufacturers protocol. DNA purity was assessed by measuring 260/280 nm absorbance ratios using a nanovue spectrophotometer (GE Healthcare Life Sciences, UK) with a threshold of 1.8 - 2.0 being set as the acceptable standard for use in WGS. Finally the integrity of the DNA was checked by agarose gel electrophoresis in the same method as described for detection of PCR products.

2.5 Whole genome sequencing (WGS)

2.5.1 DNA library preparation

DNA library preparation was undertaken using Nextera XT DNA preparation kits (Illumina, UK) and protocols. The method is as follows (and complies to the manufacturers protocol with specific variances highlighted). Each library preparation was for 24 samples.

Genomic DNA was diluted in sterile MiliQ water to a final concentration of 0.2 ng/µl. A total of 1 ng of DNA was used as the starting material. It was then simultaneously fragmented and tagged (addition of adapter sequences to the fragments) in preparation for subsequent amplification by PCR. This reaction is undertaken in a thermocycler a 55^{o}C for 5 minutes followed by holding at 10^{o}C.

The fragmented and tagged DNA was used as the template for PCR amplification. A master mix (Nextera PCR master mix; Illumina, UK) was used for this reaction. Each sample was amplified using index primers (Illumina, UK) with a unique sequence ID. This allows each sample to be identified by their unique combination of indices following pooling of all 24 samples at the final stage of the protocol. The PCR reaction was undertaken with the following programme: 72^{o}C for 3 minutes, 95^{o}C for 30 seconds followed by 12 cycles of 95^{o}C for 10 seconds, 55^{o}C for 30 seconds and 72^{o}C for 5 minutes.

50 µl of the PCR product was cleaned using 20 µl Agencourt AMPure XP beads (Beckman Coulter Life Sciences, UK), to capture PCR products of between 500bp and 100bp. This process utilises Solid Phase Reversible Immobilisation (SPRI) to capture PCR fragments of a within a selected size range, allowing removal of unincorporated primers/ primer dimers/ dNTPs/ salts and fragments outwith the desired base pair range to be removed from a low concentration DNA product.

The concentration of the PCR product was confirmed by using a Qubit HS DNA assay. For the purpose of assessing relative concentrations of DNA fragment sizes 1 µl of the PCR product for each sample was run on an Agilent nano chip using Agilent 2100 Bioanalyzer (Agilent Technologies, UK) as per manufacturers guidelines.

Following quantification of the molar concentration of the average PCR fragment size by bioanalyzer, samples were manually diluted to a concentration of 2nM. To do this 10 µl of the PCR product was diluted in 10mM Tris-Cl pH 8.5, in Tween 20 0.1%. product. The normalised DNA library samples were then pooled by mixing 5 µl of each sample and shaking on a plate mixer for 2 minutes at 1,800rpm. 5 µl of the pooled 2nM library mixture was mixed with 5 µl of 0.2 M NaOH (Sodium hydroxide). This was then vortexed briefly and incubated for 5 minutes at room temperature. Following incubation, 990 µl of pre-chilled HT1 hybridisation buffer (Nextera) was added to the pooled library. This results a 10pM denatured library in 1mM NaOH that is inverted 5 times to mix. This mixture was then placed on ice until ready to be loaded in to the sequencing cartridge (approximately 5 minutes). 600 µl of the pooled, denatured library is then loaded in to a MiSeq reagent cartridge and placed in the Illumina MiSeq sequencer.

2.5.2 Sequencing

Sequencing then proceeded using a MiSeq genome sequencer (Illumina, UK) as initially described (Bentley et al. 2008), with 500 cycles of 240bp paired end reads and minimum Phred quality score of 30. This phred quality score correlates to base call accuracy. A quality score of 30 correlates to a probability of incorrect base calls of 1 in 1,000, giving an overall base call accuracy of 99.9% as defined by (Ewing et al. 1998). Phred score is estimated by Miseq sequencer (Illumina, UK) control software.

2.6 Bioinformatic analysis of sequencing data

Sequence analysis pipelines and bioinformatics tools were accessed remotely courtesy of the Pathogen Informatics Group at the Wellcome Trust Sanger Institute, Hinxton, Cambridge to conduct this work.

2.6.1 Assembly and mapping

Fastq files from MiSeq sequencing were assembled *de novo* using Velvet (v 1.2.9) (https://www.ebi.ac.uk/~zerbino/velvet/) (Zerbino and Birney 2008). Genomes were assembled iteratively with kmer optimisation, without scaffolding and with removal of reads less than 10bp in length.

SNPs were identified using a pipeline developed at Sanger Institute by the Pathogen Genomic Group as previously described (Harris et al. 2010). In short, the various steps in the pipeline can be described as follows. The sequence reads were then mapped to an appropriate reference genome using SMALT (v 0.7.4) (http://www.sanger.ac.uk/science/tools/smalt-0) and subsequent realignment around indels using GATKv1.5.9 (McKenna et al. 2010). The default mapping parameters recommended for reads were employed, but with the minimum score required for mapping increased to 30 to make the mapping more conservative. Candidate SNPs were identified using samtools mpileup (Li et al. 2009a), with SNPs filtered to remove those at sites with a mapping depth less than 5 reads and a SNP score below 60 identify single nucleotide polymorphisms. Locations of deletions and short insertions were predicted using pindel (Ye et al. 2009) and then validated by comparing the mapping of reads spanning indels to the reference genome and to a version of the same reference with the predicted indel included. If the inclusion of the indel improved mapping, that indel was retained, and reads realigned around it as per the remapping.

The reference genomes used for this work were obtained from the European Nucleotide Archive and are listed in Table 2-3. Where an appropriate reference genome was not available for mapping on the basis of strain type or clonal background, a *de novo* assembly from an individual participant's isolate was used for mapping. From all sequenced colonies per individual the highest quality assembly was chosen for use as the 'self' reference, on the basis

of lowest number of contigs composing the assembly. The assemblies used for 'self-mapping' are included in Table 2-4.

Table 2-3 *S. aureus* reference genomes used for in this study.

Genome	Accession number	Sequence type (ST)	Clonal Complex (CC)	Reference
MSSA476	BX571857	1	1	(Holden et al. 2004)
CUHK_HK188	JFFV00000000	188	1	(Ip et al. 2014)
N315	BA000018	5	5	(Kuroda et al. 2001)
COL	CP000046	8	8	(Gill et al. 2005)
HO 5096 0412 (EMRSA15)	HE681097	22	22	(Holden et al. 2013)
MRSA252	BX571856	36	30	(Holden et al. 2004)
CA347	CP006044	45	45	(Stegger et al. 2013)
M013	CP003166	59	59	(Huang et al. 2012)

Table 2-4 *S. aureus* genome assemblies used for self-mapping.

Assembly study ID	Sequence type (ST)	Clonal complex (CC)
PSAE001_E2_IC1	188	1
PSAE002_E1_C2_24	59	59
PSAE003_E1_C3	1	1
PSAE004_E3_C5	1	1
PSAE005_E1C2	30	30
PSAE006_E3_IC2	5	5
PSAE008_E1C5	45	45
PSAE008_E2C3	123	121
PSAE009_E4C2	2867	12
PSAE010_N18C6	1	1
016-NC5	45	45
026-NC5	30	30
039-NC5	30	30
045-NC1	15	15
057-NC5	582	15
059-NC5	30	30
091-NC3	30	30
094-NC2	5	5
099-NC2	30	30
105-NC3	30	30
141-NC4	22	22
147-NC3	30	30
149-NC5	30	30
153NC4	582	15
157-NC5	30	30
250-NC4	30	30
268-NC5	2889	30
303-NC2	1	1

2.6.2 Multi-locus sequence typing

Multi-locus sequence types were extracted from genome assemblies as previously described (Croucher et al. 2011). In brief assemblies were mapped against reference alleles of the 7 loci used for MLST using BWA (Li and Durbin 2009). Mapping data was then compared to the *S. aureus* MLST database (http://www.mlst.net/) for assignment of a sequence type.

Sequence types were then assigned to their respective clonal complexes by comparison with the MLST database (http://www.mlst.net/) and on the basis of varying from founding member of the clonal complex by two allelic loci of fewer.

2.6.3 Phylogenetic reconstruction

Single nucleotide polymorphisms in the core genome were used for construction of phylogeny. To do this respective reference genomes were individually manually inspected and compared using Artemis Comparison Tool (http://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act)(Carver et al. 2005) to identify accessory regions (including plasmids, phage, antimicrobial resistance cassettes and pathogenicity islands). These accessory regions were then masked from the whole genome alignment to remove variability that could have arisen by horizontal gene transfer. Core genome SNPs were subsequently identified using snp sites (https://github.com/sanger-pathogens/snp-sites) (Page et al. 2016). Each SNP was then curated by manual inspection of the BAM files in Artemis (http://www.sanger.ac.uk/science/tools/artemis) (Rutherford et al. 2000). Any SNPs occurring in regions of low coverage or high densities of base substitutions in repetitive regions (e.g. cell wall anchored proteins such as the clumping factors or serine aspartate repeat proteins) were also subsequently removed to refine the phylogeny. The remaining core genome SNPs were

used to construct the maximum likelihood phylogeny with RAxML (v SSE3) (https://github.com/stamatak/standard-RAxML) (Stamatakis 2006). The tree was then visualised using Figtree (v 1.4.2)(http://tree.bio.ed.ac.uk/software/figtree/). The genomic regions masked from alignments are included in Tables 2-5 to 2-12.

Table 2-5 Accessory regions masked from MSSA476 reference chromosome.

Genome	Accession number	Base range	Size (bp)	Feature annotation
MSSA476	BX571857	34030..56872	22842	SCC*fus*
		427854..449454	21600	genomic island
		594406..597279	2873	*sdrC* (serine aspartate repeat protein C)
		841049..843835	2786	*clfA* (clumping factor A)
		981077..1025715	44638	prophage
		1870905..1903054	32149	genomic island
		2025476..2068037	42561	prophage
		2699750..2702467	2717	*clfB* (clumping factor B)

Table 2-6 Accessory regions masked for N315 reference chromosome.

Genome	Accession number	Base range	Size (bp)	Feature annotation
N315	BA000018	36164..89425	53261	SCC*mec* type II
		426680..427999	1319	transposon
		436159..465501	29342	pathogenicity island SaPIn2
		678300..679619	1319	transposase
		867008..873352	6344	transposase A for Tn*554*
		893906..895225	1319	transposase
		1139084..1140403	1319	transposase
		1684668..1691685	7017	transposon Tn*554*
		1761714..1763033	1319	transposase
		1839605..1839955	350	transposase
		1855378..1881633	26255	Pathogenicity island SaPIn3
		1914806..1916125	1319	transposase
		2005721..2049520	43799	Bacteriophage ϕN315
		2057159..2072272	15113	Pathogenicity island SaPIn1
		2134414..2135733	1319	transposase
		2198308..2205341	7033	transposon Tn*554*
		2389328..2389747	419	fosfomycin resistance protein *fosB*
		2670520..2677802	7282	transposon Tn*554*

Table 2-7 Accessory regions masked from COL reference chromosome.

Reference genome	Accession number	Base range	Feature annotation
COL	CP000046	34192..68085	SCC*mec* type I
		354692.. 398249	Bacteriophage L54a
		470716.. 488131	νSa-alpha
		903349.. 919266	SaPI1
		1904214.. 1941150	νSa-beta

Table 2-8 Accessory regions masked from HO 5096 0412 (EMRSA15) reference chromosome

Genome	Accession number	Base range	Size (bp)	Feature annotation
HO 5096 0412	HE681097	34163..51525	17362	SCC*mec* type IVh
		80766..83085	2319	IS element
		137306..138068	762	IS element
		141422..142156	734	IS element
		315995..318131	2136	IS element
		579077..581986	2909	*sdrC*
		818398..820577	2179	IS element
		822950..825763	2813	*clfA*
		938099..938749	650	IS element
		977509..979562	2053	IS element
		1104941..1107014	2073	IS element
		1252867..1254300	1433	IS element
		1283882..1285955	2073	IS element
		1286815..1288679	1864	IS element
		1328908..1330564	1656	IS element
		1361018..1375548	14530	ICE element
		1520143..1566314	46171	prophage
		1672402..1674168	1766	hypothetical phage protein
		1861820..1863920	2100	Arsenical resistance operon
		1890735..1892772	2037	IS element
		1902265..1903544	1279	IS element
		1943383..1945059	1676	IS element
		2021753..2029411	7658	Tn*552*-like transposon
		2042948..2087965	45017	prophage phiSa3
		2509244..2510573	1329	IS element
		2568189..2570233	2044	IS element
		2619148..2620039	891	IS element
		2631460..2633161	1701	IS element
		2700730..2701419	689	IS element
		2737148..2739841	2693	*clfB*
		2825620..2826315	695	IS element

Table 2-9 Accessory regions masked from MRSA252 reference chromosome.

Genome	Accession number	Base range	Size (bp)	Feature annotation
MRSA252	BX571856	36403..102126	65723	SCC*mec* type II
		163370..164137	767	SAR0148 hypothetical protein similar to *Borrelia burgdorferi* plasmid hypothetical protein
		409112..426412	17300	prophage
		565186..566753	1567	IS element
		616430..619150	2720	*sdrC*
		733183..765726	32543	Plasmid- *cadA/ C*, arsenical resistance operon
		884362..885170	808	SAR0838- putative membrane protein- similar to *Lactococcus lactis* plasmid pCD4 hypothetical protein
		888239..891328	3089	*clfA*
		973598..975244	1646	IS element
		998073..999020	947	IS element
		1004249..1007757	3508	IS element
		1182199..1184347	2148	IS element
		1215677..1216624	947	IS element
		1351329..1374138	22809	prophage
		1431432..1433078	1646	IS element
		1493459..1494406	947	IS element
		1592538..1636098	43560	IS element
		1636208..1637413	1205	phage intergrase
		1795317..1801661	6344	IS element
		1909322..1914672	5350	IS element
		2119976..2121622	1646	IS element
		2126786..2168433	41647	IS element
		2245178..2245861	683	IS element
		2325364..2327010	1646	IS element
		2543734..2545380	1646	IS element
		2797873..2798820	947	IS element
		2802106..2804727	2621	*clfB*

Table 2-10 Accessory regions masked from MO13 (ST59) reference chromosome.

Genome	Accession number	Base range	Size (bp)	Feature annotation
M013	CP003166	3518..4097	579	transposase
		34948..81447	46499	SCC*mec* V
		86614..88659	2045	transposase
		110199..110805	606	transposase
		138015..139781	1766	transposase
		392806..394969	2163	transposase
		515740..516469	729	transposase
		609652..611679	2027	transposase
		629524..630210	686	transposase
		865416..869547	4131	SAPI
		869671..881661	11990	SAPI
		972313..974161	1848	transposase
		1012934..1015049	2115	transposase
		1182603..1183568	965	transposase
		1533580..1575674	42094	prophage PVL
		1807619..1808762	1143	transposase
		1906194..1908789	2595	transposase
		2119135..2121206	2071	transposase
		2567189..2567810	621	transposase
		2592976..2593707	731	enterotoxin G

Table 2-11 Accessory regions masked from CUHK_HK188 (ST188) reference chromosome.

Genome	Accession number	Base range	Size (bp)	Feature annotation
CUHK_HK188	JFFV00000000	43087..59829	16742	SCC*mec*
		537069..539453	2384	*sdrC*
		785168..790121	4953	*clfA*
		815257..858400	43143	prophage
		1144669..1145661	992	transposase fragment
		1367444..1367558	114	transposase fragment
		1954850..1980718	25868	prophage
		2227732..2227818	86	transposase
		2558755..2561514	2759	*clfB*
		2647717..2658792	11075	prophage
		2678661..2691921	13260	prophage
		2699564..2739053	39489	prophage
		2765925..2790808	24883	plasmid

Table 2-12 Accessory regions masked from CA347 (ST45) reference chromosome.

Genome	Accession number	Base range	Size (bp)	Feature annotation
CA347	CP006044	36469..82717	46248	SCC*mec*
		93097..93849	752	phage integrase
		422353..424257	1904	prophage
		445210..452861	7651	transposase
		615849..618611	2762	*sdrC*
		860027..862948	2921	*clfA*
		1306168..1321644	15476	prophage
		1533066..1578893	45827	prophage
		1733623..1739967	6344	phage; erma5 and streptomycin resistance
		1897754..1898883	1129	transposase
		1902716..1915987	13271	transposase
		1974035..2004334	30299	prophage
		2011094..2016910	5816	prophage
		2083299..2114869	31570	phage; scn; chps
		2118660..2124675	6015	prophage
		2308306..2309931	1827	prophage
		2588104..2588871	767	beta-lactamase
		2751984..2754864	2880	*clfB*

2.6.4 SNP reconstruction

To calculate the number of SNPs on branches within the phylogeny as well as to identify canonical base substitutions differentiating individual isolates all polymorphic base sites were reconstructed back on to the ML tree using PAML (Yang 2007) as previously described (Harris et al. 2012). The phylogenies were visualised using Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

2.6.5 Insertion/deletion analysis

Indels within individuals colony populations, i.e. where multiple colony assessment was used, were identified from self-mapping as described in Section 2.6.1. Insertions and deletions (indels) were called as previously described, and the appropriate BAM file was then manually inspected in each indel region using Artemis to confirm identify.

2.6.6 Genome annotation

De novo assemblies in fasta format were annotated using Prokka (v 1.11) (http://www.vicbioinformatics.com/software.prokka.shtml) (Seemann 2014).

Annotation of genomes to allow inference of functional gene classes was transferred by reciprocal Fasta using a custom pearl script from reference genomes that contain functional annotation (MRSA252) (Holden et al. 2004).

2.6.7 Recombination analysis

Evidence of recombination within core genome alignments was assessed using Gubbins (http://github.com/sanger-pathogens/Gubbins). Regions of high densities of SNPs, and hence likely to have arisen from homologous recombination as opposed to horizontal gene transfer, were then masked from core genome alignments used for subsequent analysis. Outputs from Gubbins were visualised in reference to their chromosomal position and in comparison to the phylogenetic reconstruction using iCANDY (Dr Simon Harris, Wellcome Trust Sanger Institute).

2.6.8 Temporal and mutation rate analysis

Temporal analysis was used to estimate the time point at which the common ancestor of a clonal population arose. The input for this analysis was a core phylogeny constructed following removal of accessory genome components, predicted recombination and homoplastic SNPs. Linear regression analysis of root to tip genetic distance versus sampling date using TempEst (TEMPoral Exploration of Sequences and Trees (formerly Path-O-Gen)) (http://tree.bio.ed.ac.uk/software/tempest/) (Rambaut et al. 2016) was then applied.

Where linear regression analysis had been applied the mutation rate was estimated by dividing the total number of SNPs in the core genome alignment (used to produce the phylogeny) by the total number of bases in the reference chromosome multiplied by the slope of the incline as determined by TempEst analysis as above.

Mutation rate calculations for estimated duration of colonisation were based upon core genome base substitution rates for the major *S. aureus* lineages as described (Uhlemann et al. 2014a). For each reference chromosome used an expected base substitution rate per month was calculated based on the size of their respective core genomes, i.e. following removal of components listed in Tables 2-5 to 2-12. This rate was then used for estimation of time period to accumulate the level of diversity, as such the age of the colonising population. To do this the maximum pairwise distance between the two most divergent colonies in each individual was determined from the phylogenetic reconstruction. This distance was then halved on the basis of the assumption that each pair of genomes had evolved at the same rate from a common ancestor, and hence the time to the ancestor of both would be represented by this distance. The time to accumulate this calculated genetic distance was then derived using the aforementioned base substitution rate.

2.6.9 Virulence determinant profiling

To assess the genomic content of virulence determinants including toxins, resistance determinants, and adhesion genes a reference database of representative alleles for the specified genes was compiled. This was derived from an extensive review of *S. aureus* literature relating to known virulence determinants implicated in human disease, and specifically cutaneous colonisation and inflammation. Multiple reference alleles were included for detection of adhesion genes given their variability between the *S. aureus* clonal complexes.

Reference alleles used for this analysis for toxins, adhesins and antimicrobial resistance determinants used for this analysis are included in Tables 2-13 to 2-15. Candidate virulence factors were identified from previously published lists and specifically for toxins (Chen et al. 2005; McCarthy and Lindsay 2013; Thammavongsa et al. 2015), adhesion genes (McCarthy and Lindsay 2010; Foster et al. 2014), and antimicrobial resistance determinants (Holden et al. 2004; Köser et al. 2012; Aanensen et al. 2016). Sequence reads were then mapped against the compiled reference sequences using SRST2 (v 0.1.8) (https://github.com/katholt/srst2) (Inouye et al. 2014) with default of greater than 90% coverage of allele and less than 10% sequence divergence to detect the presence or absence of the determinant.

Point mutations and deletions in core genes known to confer antimicrobial resistance were separately identified by manual inspection of the BAM files. The specific genes and their associated phenotypic resistances are included in Table 2-16 and were derived from (Holden et al. 2013; Aanensen et al. 2016).

Table 2-13 *S. aureus* toxin reference alleles used for this study.

Gene	Product	Accession number
hla	Alpha haemolysin	ABD30233.1
hlb	beta-haemolysin	ABD31281.1
hld	delta haemolysin precursor	ABD31299.1
psmβ1	phenol soluble modulin beta 1	ABD30245.1
psmβ2	phenol soluble modulin beta 2	ABD30246.1
psmα1	Phenol soluble modulin alpha 1	AFD54321.1
psmα2	Phenol soluble modulin alpha 2	AFD54322.1
psmα3	Phenol soluble modulin alpha 3	AFD54323.1
psmα4	Phenol soluble modulin alpha 4	AFD54324.1
chp	chemotaxis inhibitory protein	ABD31214.1
SCIN	staphylococcal complement inhibitor	ABD30229.1
sak	staphylokinase	ABD31216.1
lukS	leukocidin S	ABD22441.1
lukF	leukocidin F	ABD21112.1
lukD	leukotoxin D	ABD31015.1
lukE	leukotoxin E	ABD31015.1
eta	exfoliative toxin A	BAF67354.1
etb	exfoliative toxin B	COW39407.1
entA	enterotoxin A	BAF68155.1
entB	enterotoxin B	AAW37877.1
entC	enterotoxin C	CAA29260.1
entD	enterotoxin D	AAB06195.1
entE	enterotoxin E	AAA26617.1
entG	enterotoxin G	CAG40902.1
entH	enterotoxin H	CAG41819.1
entI	enterotoxin I	BAB57990.4
entK	enterotoxin K	ABD22279.1
entL	enterotoxin L	BAB58170.4
entM	enterotoxin M	CAG40906.1
entP	enterotoxin P	ACZ58702.1
sen	enterotoxin N	BAB57987.4
seo	enterotoxin O	BAB57992.4
seq	enterotoxin Q	CAG43727.1
seu	enterotoxin U	AAP41901.1
ssl1	Superantigen like protein 1	CAG42156.1
ssl2	Superantigen like protein 2	CAG42157.1
ssl3	Superantigen like protein 3	CAG42158.1
ssl4	Superantigen like protein 4	CAG42159.1
ssl5	Superantigen like protein 5	CAG42160.1
ssl6	Superantigen like protein 6	CAG42161.1
ssl7	Superantigen like protein 7	CAG42162.1
ssl8	Superantigen like protein 8	CAG42163.1
ssl9	Superantigen like protein 9	CAG42164.1
ssl10	Superantigen like protein 10	CAG42165.1
ssl11	Superantigen like protein 11	CAG42169.1
tst1	toxic shock toxin	BAB58173.4

Table 2-14 Clonal complex specific *S. aureus* adhesion gene alleles used for this study.

	Strain background and accession number				
Gene	**MSSA476 (CC1)**	**N315 (CC5)**	**COL (CC8)**	**EMRSA15* (CC22)**	**MRSA252 (CC30)**
sdrC	CAG42294.1	BAB41750.1	AAW37717.1	SAEMRSA1504880	CAG39587.1
sdrD	CAG42295.1	BAB41751.3	AAW37718.1	SAEMRSA1504890	absent
sdrE	CAG42296.1	BAB41752.3	AAW37719.1	SAEMRSA1504900	CAG39588.1
fnbpA	CAG44202.1	BAB43594.1	AAW37290.1	SAEMRSA1523990	CAG41560.1
fnbpB	CAG44201.1	BAB43593.3	AAW37288.1	absent	absent
clfA	CAG42526.1	BAB41975.3	AAW36411.1	SAEMRSA1507140	CAG39851.1
clfB	CAG44333.1	BAB43728.1	AAW38650.1	SAEMRSA1525350	CAG41687.1
ebh	CAG43154.1	BAB42527.3	AAW37659.1	SAEMRSA1512970	CAG40444.1
cna	CAG44395.1	absent	absent	SAEMRSA1525930	CAG41749.1
eap	CAG43667.1	BAB42081.1	AAW36971.1	SAEMRSA1508100	CAG41016.1
sraP	CAG44357.1	CAG44357.1	AAW38674.1	SAEMRSA1525590	CAG41711.1
isdA	CAG42838	BAB42226.1	AAW38019	SAEMRSA1509590	CAG40105
isdB	CAG42837.1	BAB42225.1	AAW38018	SAEMRSA1509580	CAG40105
isdH	CAG43459	absent	AAW38309	absent	absent
sasC	CAG43485	BAB42845.1	AAW38334	SAEMRSA1516640	CAG40832.1
sasD	CAG41876.1	BAB41349.3	AAW38762.1	absent	CAG39163.1
sasG	CAG44197.1	BAB43588.3	AAW37284.1	SAEMRSA1523950	absent

Gene products: sdr- serine aspartate repeat protein; fnbp- fibronectin binding protein; clf- clumping factor; ebh- extracellular matrix binding protein; cna- collagen adhesion; eap- extracellular adherence protein; sraP- serine rich adhesin for platelets; isd- iron regulated surface protein; sas- *S. aureus* surface protein. * Accession numbers not available for individual gene sequences- locus ID within reference genome (accession number: HE681097) is provided as an alternative.

Table 2-15 Antimicrobial resistance determinant alleles used in this study.

Gene	Product	Resistance	Accession number
blaZ	Beta-lactamase	Penicillin	CAG40822.1
blaZ-D	Beta-lactamase	Penicillin	AAQ84127.1
blaZ-B	Beta-lactamase	Penicillin	AAC35853.1
blaZ-A	Beta-lactamase	Penicillin	AAA26644.1
mecA	Penicillin binding protein 2a	Methicillin	BAA86650.2
ermA	rRNA adenine N-6-methyltransferase	Macrolides	BAA82205.2
ermB	rRNA adenine N-6-methyltransferase	Macrolides	ABQ00061.1
ermC	rRNA adenine N-6-methyltransferase	Macrolides	AAA20192.1
tetK	Tetracycline efflux protein	Tetracyclines	AAA19179.1
tetM	Tetracycline resistance protein	Tetracyclines	BAB56560.4
linA	Lincomycin resistance protein LinA	Clindamycin	AAS50177.1
catA1	Chloramphenicol acetyltransferase	Chloramphenicol	CAA26105.1
catA2	Chloramphenicol acetyltransferase	Chloramphenicol	AAQ55242.1
catA3	Chloramphenicol acetyltransferase	Chloramphenicol	CAA24586.1
fexA	Chloramphenicol/ florfenicol exporter	Chloramphenicol	AGJ70569.1
fusB	Fusidic acid resistance protein	Fusidic acid	AY047358.1
fusC	Fusidic acid resistance protein	Fusidic acid	CAG41812.1
ileS-2	Isoleucine tRNA transferase	Mupirocin	CAA53189.1
str	Streptomycin resistance protein	Streptomycin	CAA29839.1
aadD(9)	Streptomycin 3"-adenylyltransferase	Streptomycin	BAA82204.2
aphA-3	Aminoglycoside phosphotransferase	Kanamycin	BAF82029.1
aadD	Kanamycin nucleotidyltransferase	Kanamycin	CAA27142.1
qacA	Antiseptic resistance protein	Antiseptics	BAB47540.1
qacB	Antiseptic efflux protein QacB	Antiseptics	AAQ10692.1
qacC	Quaternary ammonium compound resistance protein QacC	Antiseptics	AAA26666.1
qacJ	QacJ protein	Antiseptics	CAD55144.1
qacG	Quaternary ammonium compound resistance protein QacG	Antiseptics	CAA76542.1
qacH	Quaternary ammonium compound resistance protein QacH	Antiseptics	CAA76544.1
copA	Copper-exporting P-type ATPase A	Copper	CAG41616.1
mco	Multicopper oxidase	Cooper	CAG39732.1
arsB	Arsenical pump membrane protein	Arsenic	AAA25637.1
arsC	Arsenate reductase	Arsenic	AAA25638.1
cadA	Cadmium transporting ATPase	Cadmium	CAG39734.1
cadD	Cadmium resistance protein	Cadmium	BAB43874.1

Table 2-16 Point mutations conferring antimicrobial resistance assessed for in this study.

Gene	Product	Resistance	Substitution/ mutation	Reference
gyrB	DNA gyrase subunit B	Quinolones	D437N	(Griggs et al. 2003)
gyrA	DNA gyrase subunit A	Quinolones	E88K, S84L	(Griggs et al. 2003)
grlA	DNA topoisomerase IV subunit A	Quinolones	S80F, S80Y	(Griggs et al. 2003)
rpoB	DNA directed RNA polymerase subunit beta	Rifampicin	S464P, Q468R, D471Y, A477V, A477D, H481Y, R484H, D550G	(Aubry-Damon et al. 1998)
fusA	Translation elongation factor	Fusidic acid	V90I, E444K, G451V, M453I, H457Y, L461K, L461S, P478S, M651I	(Chen et al. 2010)
			P406L, G452S	
			H457Q	(Lannergård et al. 2008)
				(Castanheira et al. 2010)
rplf	Ribosomal protein L6	Fusidic acid	Stop at nt 229 Stop at nt 241 Deletion of nt 203-231 Deletion of nt 404-427	(Norström et al. 2007)
ileS-1	Isoleucyl tRNA transferase	Mupirocin	G593V, V631F, V588F	(Hurdle et al. 2005)
dfrA	Dihydrofolate reductase	Trimethoprim	F99Y, F99S, L41F, H150R	(Vickers et al. 2009)

2.6.10 Functional gene classification

Genes affected by non-synonymous mutations were classified according into functional groups on the basis of defined or predicted function of the protein coding sequences. This classification was based upon the original functional classification scheme from the *Escherichia coli* genome (Serres et al. 2001) and amended for its application to Gram-positive organisms (Castillo-Ramírez et al. 2011; Harris et al. 2015; Weinert et al. 2015).

2.6.11 Pangenome analysis

In order interrogate gene content across the whole genome, in isolate populations derived from disease cases and controls, the annotated assemblies were used to construct a pangenome with ROARY (https://sanger-pathogens.github.io/Roary/) (Page et al. 2015).

2.6.12 Discriminant analysis of the principle components (DAPC)

The output of gene content from ROARY analysis was converted to a binary matrix of presence or absence of gene/ homology group in excel. DAPC (Jombart et al. 2010) was then applied using the R package adegenet (Jombart 2008) to assess if the isolate genotype could be differentiated on the basis of disease status as previously described (Weinert et al. 2015).

2.7 Statistical analysis

Formal sample size requirement estimates were not made prior to the comparative studies presented in Chapter 4, 5 and 6. Sample sizes were selected on the basis of previous studies and what was practicable. Given the exploratory nature of studies presented in Chapters 4 and 5 stochastic statistical testing was also not necessarily appropriate.

All statistical tests were carried out as two tailed tests. Stochastic statistical tests are based on unpaired comparisons using methods based on the Chi-square tests or, where individual cell values were small, Fisher's exact test, with a fraction added to each cell (small value adjustment) when necessary. The statistical analyses included in Chapter 5, Section 5.8.2 were performed in collaboration with Dr Alison Mather, The Wellcome Trust Sanger Institute. This included Fishers exact test and Chi2 comparison, using R version 3.3.1 R Core Team (2016). R:

A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Analyses presented in Chapter 6 were performed with Dr Robert Dawe, Honorary Reader in Dermatology, Ninewells Hospital and Medical School. This was undertaken using algorithms within Stata 14.2, Stata Corps, Texas, 2016. The comparison of unpaired proportions presented in Chapter 6 were derived from a modified Chi-2 test using a method based upon that described by Newcombe and Altman (Altman 2000). To aid interpretation of the relevance of differences in chapter 6, 95% confidence intervals for relevant differences, not only P- values were presented. Adjustments for multiple comparisons were not made. The significance threshold taken for all analyses was set at 0.05, but actual P values and associated confidence intervals were assessed (so not inappropriately interpreting almost identical P values (such as P=0.005 and P= 0.045) as if they were markedly different.

3 Genomic investigation of the emergence of the first Methicillin resistant *Staphylococcus aureus* (MRSA)

3.1 Introduction

The introduction of penicillin in the 1940s revolutionised the treatment of Staphylococcal infection in a time when systemic infection had reported mortality rates exceeding 80% (Skinner and Keefer 1941). However within four years penicillin resistant strains were being reported (Kirby 1944), increasing in incidence initially nosocomially (Barber and Rozwadowska-Dowzenko 1948) finally becoming pandemic by mid-1950s (Rountree 1955). To circumvent this growing resistance the semi-synthetic β-lactam, Methicillin (Celbenin), was introduced in the UK in 1959 (Knox 1960). Less than a year later, a screening study at the Staphylococcal Reference Laboratory at Public Health England observed the first methicillin resistant *S. aureus* (Jevons 1961). Screening of more than 5,440 strains, obtained during the period of Celbenin's introduction (October 1959- November 1960), led to the identification of 3 resistant isolates. Further investigation revealed the isolates shared a common phage type and resistance profile (Penicillin, Streptomycin, Tetracycline, Methicillin) (Jevons 1961). All three isolates were noted to originate from the same south London hospital where Methicillin had been used clinically in a single instance, but with none of the affected individuals having been treated with the drug. Two years later instances of invasive MRSA of the same phage type were being reported in Denmark (Eriksen and Erichsen 1964) heralding the spread of the first epidemic MRSA clone.

These early MRSA isolates are known to have arisen in *S. aureus* CC8, with ST250 being the initial recipient background of SCC*mec*. Whilst being a successful lineage in Europe during the 1960s and 70s, incidence gradually faded by the 1980s, and it is now rarely reported (Enright et al. 2002; Monecke et al. 2011). The single locus variant of ST250, ST247, became a seemingly more successful strain, spreading across Europe as well as America but its prevalence has diminished since the late 1990s (Roberts et al. 1998). Now more than fifty years since the first appearance of MRSA, multiple MRSA lineages have emerged in genetically

diverse backgrounds and successfully spread. Niche adapted clones, including healthcare, community and livestock associated MRSAs, are now prevalent and well established worldwide.

3.2 Aims and Objectives

Previous epidemiological evidence has always suggested that methicillin resistance arose around the period of the drugs introduction, and as an adaptive response to the therapy (Jevons 1961). As such the time point at which *S. aureus* first acquired the SCC*mec* element conferring methicillin resistance, has never been determined. Whole genome sequencing of a collection of the very first MRSA isolates derived from 1960 through to 1989 was undertaken. The aim of this work was to use phylogenetic reconstruction this population to trace the evolutionary history of the archetypal MRSA lineage. Temporal analysis was applied to estimate the age of the strain background, and as such the time point at which the SCC*mec* was first acquired. In addition this work aimed to assess the genetic content of the population to understand the impact of antimicrobial usage on the organism during this historical period.

The WGS data used for this analysis are from a historical collection of MRSA samples and have been sequenced as collaborative study. Isolates were kindly provided by Public Health England, the Statens Serum Institute and Rockefeller University and whole genome sequencing was undertaken at the Wellcome Trust Sanger Institute, Hinxton, Cambridge.

3.3 Origins of the Historic MRSA collection

The isolates utilised for this study are derived from the culture collections of the Staphylococcal reference laboratories at Public Health England (PHE), Colindale and the Statens Serum Institute (SSI) in Copenhagen. They represent the first documented isolates of MRSA identified anywhere in the world. These isolates were collected in Europe,

predominantly in the UK and Denmark, between 1960 and 1989. A total of 209 CC8 isolates were sequenced for this study. Of these 188 originate from the PHE collection and are principally from the UK, collected between 1960 and the late 1970s. The remaining 21 isolates are from the SSI collection, isolated in Denmark between 1964 and 1989. Included within the PHE collection are the 8 isolates from the original study identifying MRSA (Jevons 1961).

(Isolate identifiers, year of isolation and country of origin are available for the Historic collection in online supplementary data file via https://figshare.com/ ;Chapter 3- Historic MRSA suppl. Data)

3.4 Population structure of the historic collection

From the assembled genomes the multi-locus sequence type was extracted (as per method Section 2.6.2) and revealed that two STs predominated, ST250 (n=126; 60.3%) and ST247 (n=78; 37.3%). The remaining 5 (2.4%) isolates were allelic variants including, 4 single locus variants (SLV) of ST247 (ST335 (n=3), ST3578 (n=1)) and one triple locus variant (TLV) of ST250. This TLV (ST3526) was derived from the PHE collection with a date of origin of 1967 and was subsequently retained as an out-group for the remainder of the analysis.

To characterise the population structure of the collection and construct a phylogeny the sequence reads of the 209 isolates were mapped to the chromosome of the CC8 reference *S. aureus* COL (accession number: CP000046) (Gill et al. 2005). This reference is notable as it is also derived from this early MRSA lineage, and was first identified by the Staphylococcal Reference Lab at PHE in the 1960s (Dyke et al. 1966). To construct the phylogeny on the basis of vertically inherited SNPs, mobile genetic elements were excluded from the alignment. Regions of high densities of base substitutions within the remaining core genome alignment were identified using Gubbins (as described in Section 2.6.7), and these predicted regions of homologous recombination were additionally removed (Table 3-1).

Table 3-1 Predicted regions of recombination from Gubbins analysis. The following regions were excluded from the core genome alignment used for construction of phylogeny. Base positions are relative to *S. aureus* COL reference (accession number: CP000046)

Base range	Size (bp)	Region/ gene
107826..107876	50	SACOL0095 *spa*
148032..148517	485	SACOL0132 hypothetical protein
247291..247353	62	SACOL0209 Staphylocoagulase precursor
324406..331320	6914	Between SACOL0283 and SACOL0294- hypothetical proteins
638169..638815	646	SACOL0609 *sdrD*
639229..639274	45	SACOL0609 *sdrD*
639352..639547	195	SACOL0609 *sdrD*
863308..863375	67	SACOL0836 hypothetical protein
882663..882802	139	SACOL0856 *clfA*
882877..882901	24	SACOL0856 *clfA*
1069229..1069250	21	SACOL1062 Bifunctional autolysin
1070920..1070929	9	SACOL1062 Bifunctional autolysin
1178201..1178292	91	SACOL1170 hypothetical protein
1182246..1182272	26	SACOL1176 hypothetical protein
1182359..1182384	25	SACOL1177 hypothetical protein
1482151..1482261	110	SACOL1472 *ebH*
1736804..1736834	30	SACOL1707 *radC*
1823478..1823508	30	SACOL1781 *sasI*
1878411..1880350	1939	SACOL1822/1823/1824 Arsenical resistance operon
1965970..1966048	78	SACOL1910 hypothetical protein
1974652..1974684	32	SACOL_tRNA-Pseudo-1
2008839..2008852	13	Between SACOL1946 and SACOL1947
2075427..2076036	609	SACOL2015- integrase/ recombinase
2079662..2079678	16	SACOL0219 *sdrH*
2201861..2201928	67	Frameshift of SACOL2141 site specific recombinase family protein
2570133..2570229	96	SACOL2511 *fnbA*
2571423..2571429	6	SACOL2511 *fnbA*
2616733..2616763	30	Between SACOL2557 and SACOL2558
2711513..2711783	270	SACOL2652 *clfB*
2711854..2711948	94	SACOL2652 *clfB*

A total of 4220 core genome SNP sites were then used to construct the maximum likelihood phylogeny rooted with NCTC8325 (Figure 3-1). Evident from the phylogeny of the whole collection is a diverse population with multiple clades. Closely related isolates derived from the UK are seen to cluster together, whilst the Danish isolates fall predominantly within 3 main clusters distributed across the population. There is also clear separation of the population on the basis of ST, with ST250 isolates forming the most basal part of the population. The ST247 isolates extend out from this ST250 population forming their own distinctive clade. Also

notable is that isolates within the ST247 sub-population account for the majority of the contemporaneous parts of the population, with dates of origin ranging from 1964 through to 1989 (Figure 3-3; Chapter 3-Historic MRSA Suppl. data via figshare).

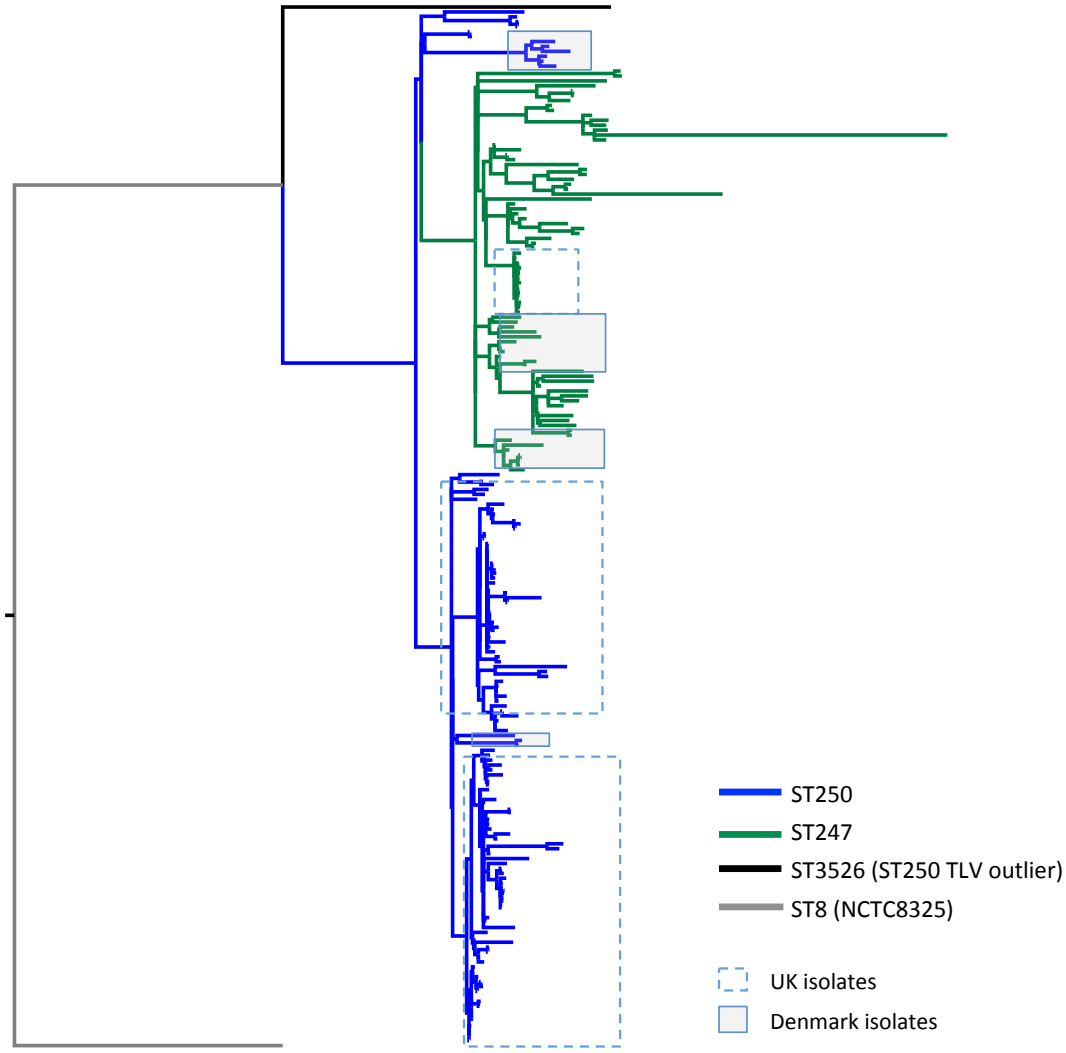

Figure 3-1 Population structure of the historic collection. Maximum likelihood core phylogeny of the 209 CC8 isolates. Branches are coloured according to multi locus sequence type – ST250 (blue), ST247 (green), ST250 TLV (triple locus variant) (black) and ST8 (grey). Boxes around taxa indicate their country of origin- either UK (perforated box) or Denmark (shaded box). Tree is rooted using ST8 reference (NCTC8325) as the inferred ancestral background of the ST250 lineage.

Illustrated in the expanded view of the core phylogeny in Figure 3-2 are the Jevons isolates (Jevons 1961). These samples were collected from three individuals during the period between July and November 1960 from a single hospital in the South London area. The sources of each of the isolates as well as the resistance determinants are shown in Table 3-2. Demonstrated by the phylogeny is the close relatedness of these 8 isolates, separated by only 6 core genome SNPs. Their relative genetic distances combined with their position in the phylogeny and the epidemiological evidence suggest that cross transmission could have occurred between these individuals (Patients A, C and Nurse B). These samples are also found to cluster with an additional 6 isolates derived from the PHE collection, three of which are known to originate from the same region of London, and within a year of the Jevons study isolates. Full epidemiological data are not available for these isolates, but their genetic relatedness and region of origin are indicative that this MRSA clone was distributed within this geographical area, and therefore are likely representative of local transmission.

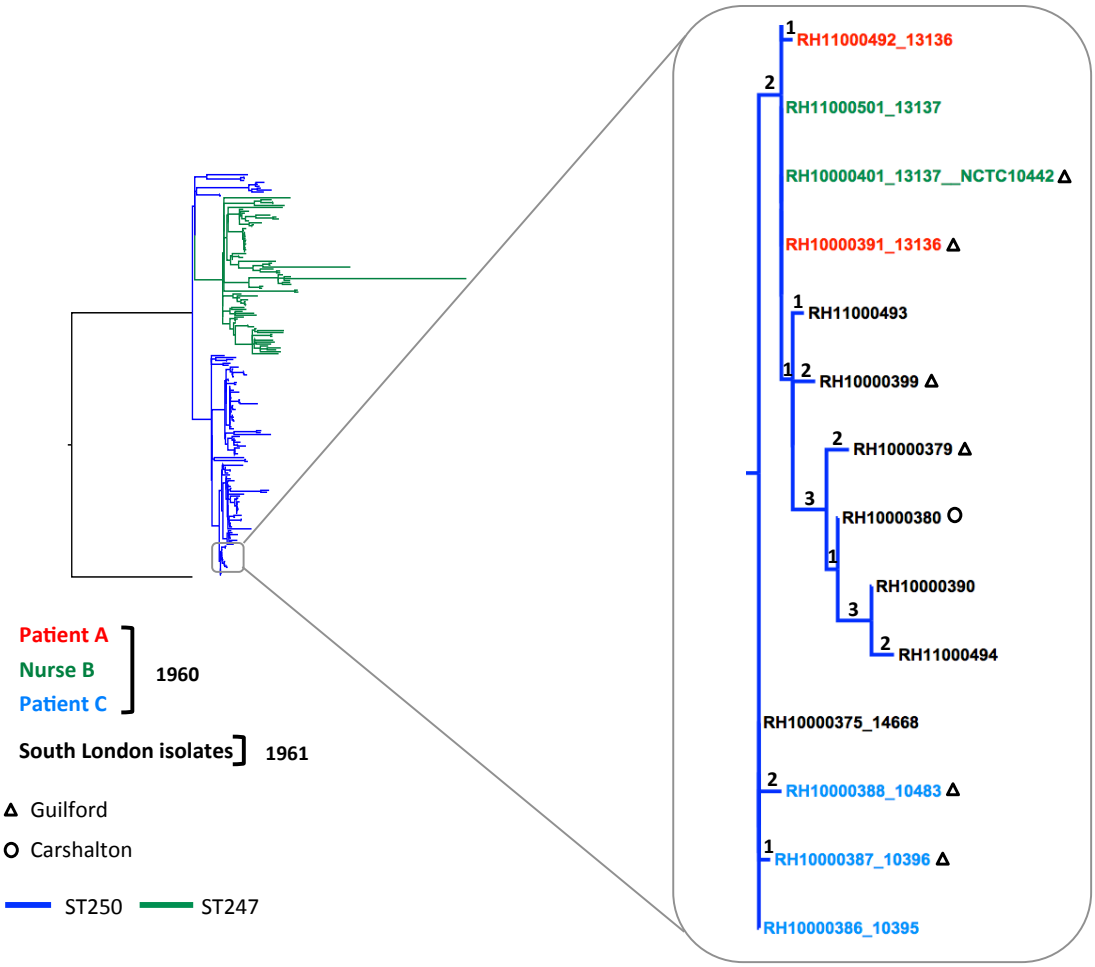

Figure 3-2 Jevons Isolates within the Historic collection. Expanded view of maximum likelihood core tree showing the 8 isolates described by Patricia Jevons in 1961. Numbers superimposed above branches correlate to SNP distances between isolates. Coloured taxon IDs (blue/ red and green) correlate to the individual the sample was derived from in the 1961 description. Isolate IDs are known to have derived from South London region or presumptively given genetic distance. Shapes indicate the South London region that the isolates originate from, where the epidemiological data was available. Year of isolation of isolate is indicated next to taxon ID colour scheme. Colouring of branches (blue) indicates isolates are all of ST250 background.

Table 3-2 Minimum inhibitory concentration (MIC) to Methicillin (Celbenin) derived from the original Jevons description, BMJ 1961. MIC values represent the variation noted between colonies. The expected range of sensitivity to Celbenin in coagulase positive staphylococci is 1.25-2.5µg/ml.

Jevons strain ID	PHE ID	Celbenin MIC (µg/ml)	Source	Date of Isolation	Resistance determinants present
13137	RH11000501	12.5-25	Nurse B	02.10.60	*blaZ* *mecA* *tetK*
13137	RH1100040 (NCTC 10442)	12.5-25	Nurse B	02.10.60	*blaZ* *mecA* *tetK*
13136	RH10000391	6.5-12.5	Patient A	02.10.60	*blaZ* *mecA* *tetK*
13136	RH11000492	6.5-12.5	Patient A	02.10.60	*blaZ* *mecA* *tetK*
10395	RH10000386	6.25	Patient C	21.07.60	*blaZ* *mecA* *tetK*
10396	RH10000387	3.125	Patient C	05.07.60	*blaZ* *mecA* *tetK*
14083	RH10000388	1.6	Patient C	28.10.60	*blaZ* *tetK*
14668	RH10000375	25	Patient C	08.11.60	*blaZ* *mecA* *tetK*

3.5 Genetic basis of antimicrobial resistance within the population

3.5.1 Resistance genotype of the population

The initial description of this archetypal MRSA lineage described multiple resistances in the strain background in addition to methicillin. These included penicillin, tetracycline and streptomycin. Analysis of the whole genome data revealed resistance to multiple classes of antimicrobials reflecting the antibiotics in use at the time (Figure 3-3). The methicillin resistance determinant, *mecA*, was present in 197 (94%) of the isolates, while carriage of the beta-lactamase gene, *blaZ*, was found in 180 (86%) isolates. From the Jevons study a single colony isolate from Patient C was sensitive to methicillin (Celbenin) (Table 3-2: Jevons study ID 14083/ PHE ID RH10000388), and analysis of this isolate showed that *mecA* was absent. Tetracycline resistance is prominent within the collection, with 187 (89.4%) of the population carrying *tetK* including all of the Jevons isolates. Additionally 94 (44.9%) of the isolates were carrying the alternative tetracycline resistance determinant, *tetM*, which was not identified in

any of the Jevons isolates. In total 94 (44.9%) of the isolates were carrying both *tetK* and *tetM* which has been shown to confer higher levels of resistance to tetracycline (Trzcinski et al. 2000). Despite their described phenotypic resistance no horizontally acquired streptomycin resistance determinants were identified in the Jevons isolates. Otherwise within the population the streptomycin resistance determinants *aad9* and *str* were found in 39.7% (n=83) and 0.47% (n=1) of the isolates respectively.

Resistance determinants to numerous other antimicrobial agents were identified within the collection including: erythromycin (*ermA*) in 83 (39.7%) isolates, chloramphenicol (*catA1, catA2 and catA3*) in 4 (1.9%), 16 (7.6%), and 1 (0.47%) isolates respectively. Kanamycin resistance determinants *aphA3* and *aadD* were found in 16 (7.6%) and 1 (0.47%) isolates respectively. In addition to the acquired resistance genes, a point mutation associated with resistance to novobiocin (*gyrB*, R144I) was found in 17 (8.1%) of isolates, and mutations conferring resistance to fusidic acid (*fusA*, P404L) and trimethoprim (*dfrA*, F99Y) were found in a single instance.

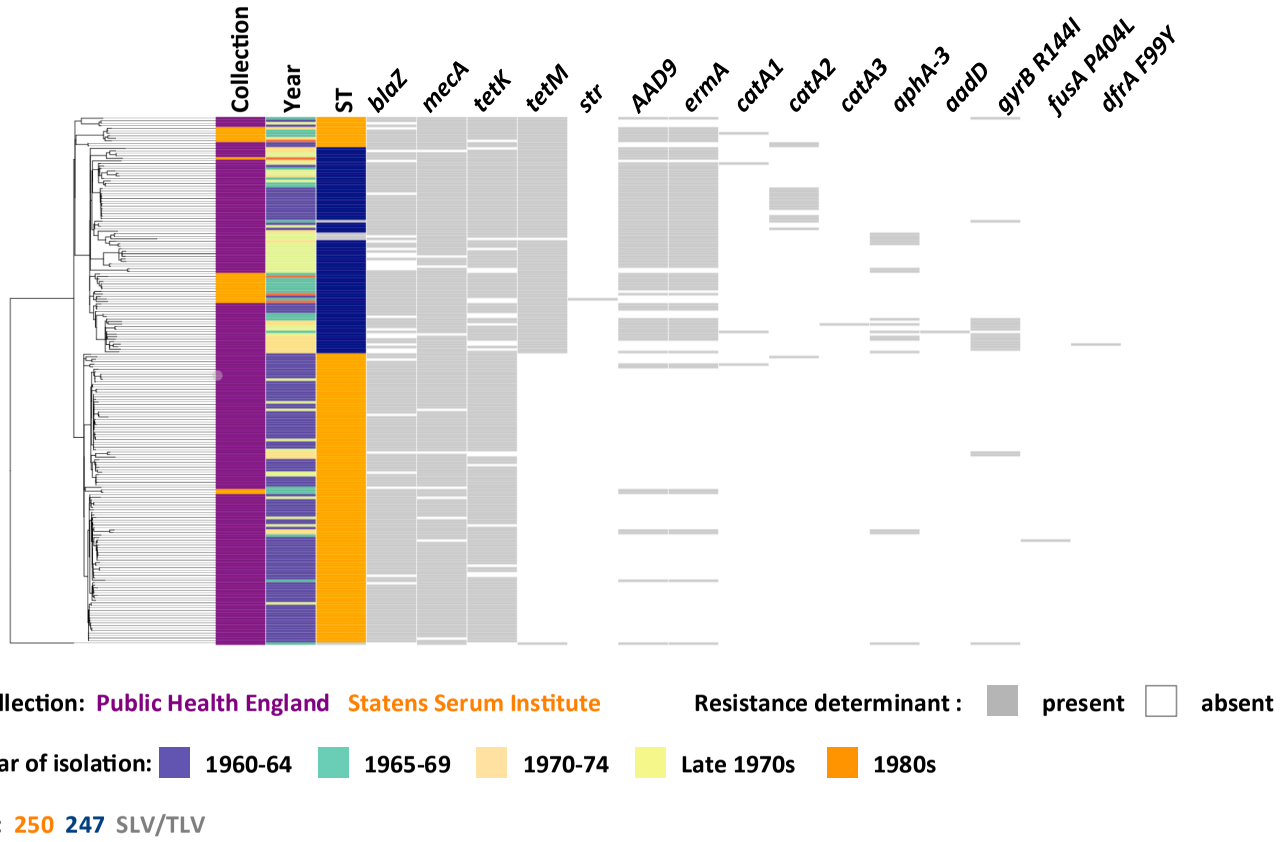

Figure 3-3 Resistance genotype of the historic collection. Presence/ absence of resistance determinants across the collection of all 209 isolates in comparison with their distribution throughout the phylogeny. Collection refers to their reference laboratory (and presumed country of origin) PHE (UK- purple) and SSI (Denmark- orange). Year of isolation is coloured in accordance to date ranges. Sequence types are displayed as orange (ST250), blue (ST247) and grey (allelic variants). Grey shading beneath resistance determinants indicates presence, and white indicates absence. . (Gene abbreviations are as defined in Materials and Methods section 2.6.9, Table 2-15).

3.5.2 The SCC*mec* element in the historic collection

Methicillin resistance as the defining feature of this lineage is the result of carriage of *mecA* on the SCC*mec* cassette. Having assessed the prevalence of the methicillin resistance marker, further detailed analysis was then undertaken to assess the types of SCC*mec* within the collection.

Previously published studies have demonstrated that this archetypal clone carried a type I SCC*mec* element, which was the first of these mobile elements to be characterised (Katayama et al. 2000; Ito et al. 2001). The first description of the type I element was based upon the SCC*mec* from *S. aureus* strain NCTC10442 (Ito et al. 2001) identified in 1960 in the Jevons study (Table 3-2) and is included in this collection. The only resistance determinant carried on this MGE is *mecA*, which combined with a truncated variant of the gene encoding the MecR1 regulatory protein is known as a class B *mec* gene complex. This complex with type 1 chromosomal recombinases (*ccrA1* and *ccrB1*) define this sub-type of the element (IWG-SCC 2009). Identified during the classification of the type I SCC*mec* was the presence of a frame shift mutation affecting *ccrB1* (Ito et al. 2001). Comparison of mapping data derived from 193 of the isolates in the collection with an intact SCC*mec* element, 192 were of type I and had NCTC10442-type *ccrB1* frame shift mutation. A single isolate was found to be carrying alternative type recombinases which subsequent comparison to the known SCC*mec* variants led to its identification as a type IVh element. This isolate is interestingly the ST250 TLV (ID: RH12000692_7401696; online supplementary data file Chapter 3-Historic MRSA Suppl. data) isolated in 1967. The remaining 16 isolates were either not carrying *mecA* (n=12) or had incomplete elements with internal deletions (n=4).

3.6 Investigating the evolutionary events leading to emergence of the first MRSA

3.6.1 The acquisition of SCC*mec*

In order to reconstruct the evolutionary history of MRSA in the population and specifically determine if SCC*mec* entered this *S. aureus* population once or on multiple occasions, variation within this MGE was then assessed.

In order to do this a phylogeny was constructed on the basis of SNPs within the type I SCC*mec* element. Inspection of the phylogeny and distribution of SNPs suggested that some of this variation was the result of homologous recombination. Subsequent Gubbins analysis revealed there were two predominant regions of recombination within the MGE as shown in Figure 3-4. Within the 192 type I SCC*mec* elements analysed, a total of 194 SNP sites were present in the regions of recombination identified by Gubbins. These included: *pls* (methicillin resistant surface protein) and an intergenic region between SACOL0030 (predicted hypothetical protein) and SACOL0031 (glycerophosphoryl diester phosphodiesterase). A total of 124 of these 194 SNPs were within the *pls* LPXTG surface protein (length 4764bp), and 31 SNPs were in the 549bp intergenic region between SACOL0030 and SACOL0031. Exclusion of these regions of recombination left a total of 39 SNPs differentiating the 192 elements. A further single base substitution within the IS*431* transposase was excluded from the alignment. Manual inspection of the mapping of this region revealed that it was present in multi-copy within the SCC*mec,* and there was SNP heterogeneity in this position therefore making it unsuitable for accurate phylogenetic reconstruction.

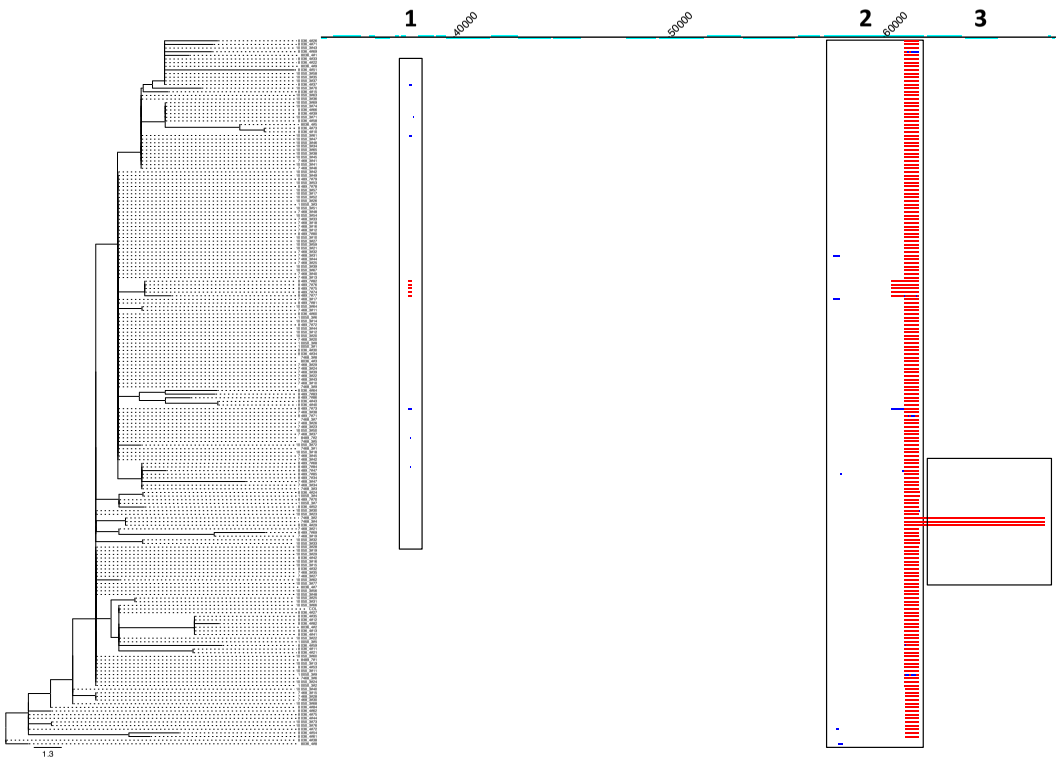

Figure 3-4 Recombination within SCC*mec* type I elements of the historic collection. Regions of variation arising in 192 type I elements. Red blocks indicate regions of recombination. Isolates are ordered according to the phylogenetic tree displayed on the left. The relative chromosomal position in accordance to *S. aureus* COL (accession: CP000046) where recombination has arisen is displayed along the top of the figure, where protein-coding sequences (CDS) are indicated in light blue. Numbers in bold indicate the specific regions of recombination as follows: **1**- Intergenic region between SACOL0030 (hypothetical protein) and SACOL0031 (glycerophosphoryl diester phosphodiesterase). **2**-SACOL0052 (*pls* - Methicillin resistant surface protein). **3**- Across SACOL0051 (hypothetical protein)/ SACOL0052 (glycosyl transferase group 1 family protein) and SACOL0054 (pseudogene).

The subsequent phylogeny (based on 38 core SNP sites) revealed 116 of the isolates to be carrying identical SCC*mec* elements, including all 7 of the *mecA* positive Jevons isolates (Figure 3-5; Table 3-2). The remainder carried elements with a maximum SNP distance of 8 differentiating them. This limited SNP distance as well as the phylogeny of the elements demonstrate the close relationship of the SCC*mec* within this historic population, and suggests they derived from a common ancestral element.

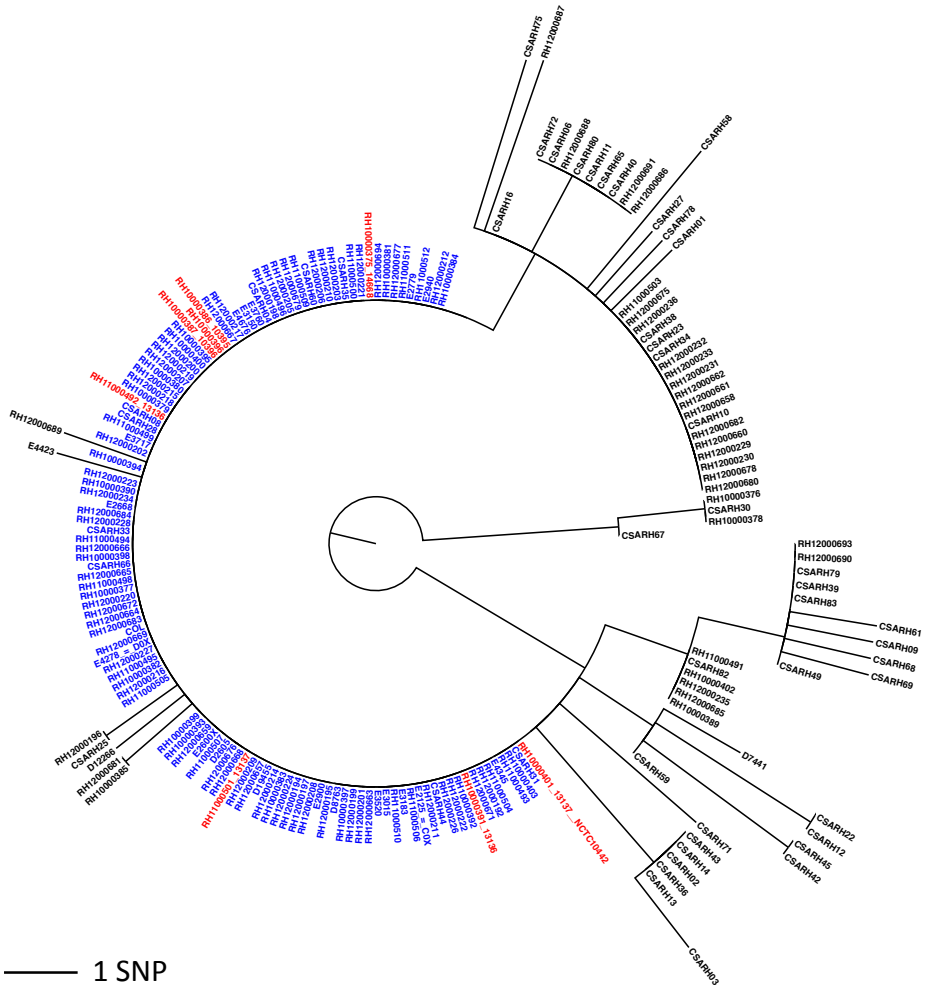

Figure 3-5 Maximum likelihood core SNPs tree SCC*mec* type I elements in historic collection. A total of 192 isolates used to construct the tree. Highlighted in blue are isolates with genetically identical SCC*mec* following removal of predicted recombination events within the MGE. Jevons isolates are arbitrarily distributed throughout the phylogeny, as they are genetically identical (highlighted in red). *S. aureus* COL is included as a reference and is a member of the basal/ identical population. Scale bar indicates SNP distance.

3.6.2 Estimating the age of the Archaic MRSA clone

In order to investigate the temporal effects surrounding the evolution of the archaic MRSA clone a subset of the PHE collection with an exact date of isolation were selected. A total of 122 samples, isolated between 1960 and 1968 formed the basis of this analysis. This sub-group was chosen specifically as there was intact epidemiological information regarding the isolates origin including year of isolation and country of origin.

A core phylogeny was produced for this subset. To ensure that the temporal analysis was conducted on vertically inherited SNPs all variation having arisen through gain of MGEs as well as recombination was masked from the alignment. In addition a total of 59 homoplastic SNPs (Table 3-3; presented at the end of section 3.6.2) were excluded to remove heterogeneity having arisen as the result of strong selective pressure. The remaining 1096 core genome SNPs were used for construction of the phylogeny. Regression of the root to tip distance versus sampling date with TempEst was then used to infer the time point at which the common ancestor of these isolate arose. As shown by the X-axis intercept in Figure 3-6 this suggested that the strain background of this ST250 population arose in 1946, almost fourteen years prior to the first reports of clinical use of methicillin (Knox 1960; Rolinson et al. 1960; Stewart et al. 1960).

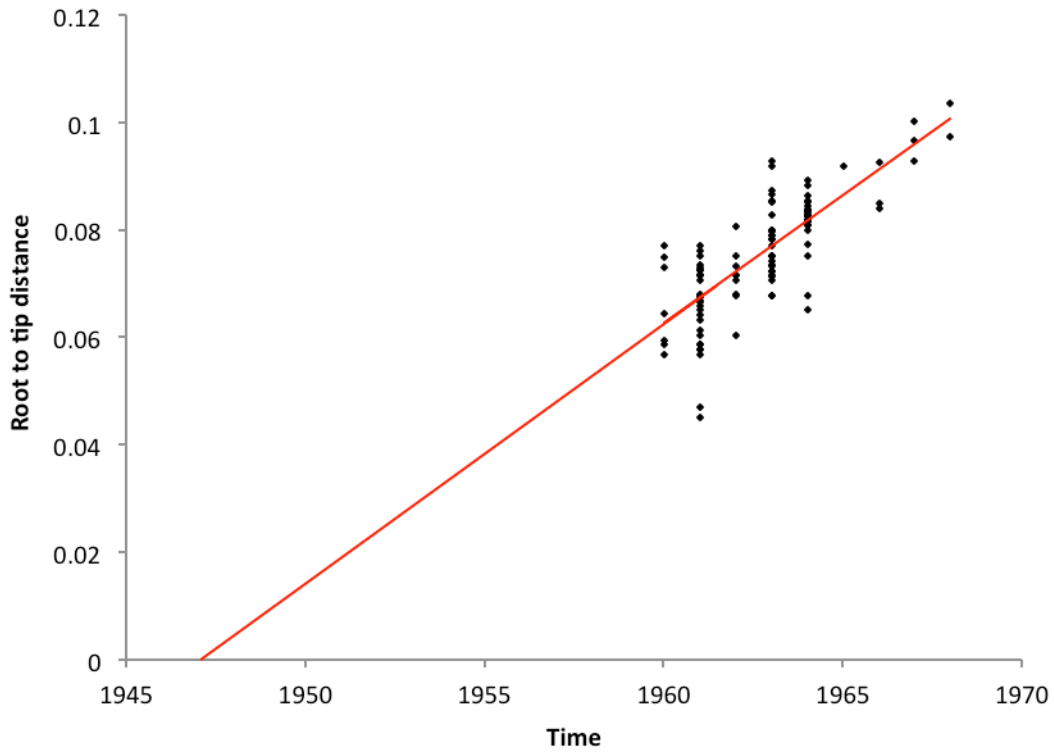

Figure 3-6 Root to tip distance regression analysis using TempEst. Core genome SNPs within 122 of the Public Health England derived historical collection isolates were used. All MGEs, predicted recombination and homoplasies were removed from alignment to produce the maximum likelihood tree used as the input for this analysis. The vertical axis represents genetic distance and the horizontal axis represents time, with the X-axis intercept year being 1946, correlation co-efficient: 0.7907, R^2: 0.6525, slope rate: 4.7424E-3 and residual mean: 4.0787E-5.

As a further assessment of the micro evolutionary history of these isolates a mutation rate was calculated. To do this total core genome SNPs in this subpopulation (n=1096) were divided by the total number of bases in the reference genome used for mapping (COL accession number: CP000046; 2,089,422 bases) multiplied by the slope incline (rate) from the TempEst output 4.7424E-3 (Figure 3-6). This gave an estimated mutation rate of $1.8x10^{-6}$ SNPs/site/year, corresponding to approximately 3.7 SNPs per year. This substitution rate falls within the reported ranges of multiple successful *S. aureus* lineages (Uhlemann et al. 2014a) and therefore it does not appear that the long-term storage, or repeated culture passage of the isolates has affected this observation.

Table 3-3 Homoplastic SNPs identified in core genome alignment of 122 isolates used for temporal analysis. Base positions are relative to *S. aureus* COL reference (accession number: CP000046). NS: non-synonymous, S: synonymous, I: intergenic.

Base position	SNP type	Base change	Region/ gene
5491	NS	A->T	SACOL0005 *gyrB*
108948	I	C->A	Between SACOL0095 *spa* and SACOL0096 *sarS*
113598	I	G->T	Between SACOL0099 *sirA* and SACOL0100 cysteine synthase
148687	I	T->C	SACOL0132 hypothetical protein
354685	S	A->G	SACOL0317 Lipase precursor
469173	NS	C->T	SACOL0467 conserved hypothetical protein
615980	NS	G->T	SACOL0593 *fusA*
617225	NS	C->A	SACOL0593 *fusA*
756666	NS	T->A	SACOL0732 hypothetical protein
756735	I	T->A	Between SACOL0732 hypothetical protein and SACOL0733 sugar efflux transporter
809620	NS	T->A	SACOL0786 hypothetical protein
809621	NS	T->C	SACOL0786 hypothetical protein
809622	NS	C->G	SACOL0786 hypothetical protein
809676	NS	A->T	SACOL0786 hypothetical protein
809677	NS	C->T	SACOL0786 hypothetical protein
809678	NS	G->C	SACOL0786 hypothetical protein
809706	S	G->C	SACOL0786 hypothetical protein
883048	S	C->T	SACOL0856 *clfA*
883102	NS	T->A	SACOL0856 *clfA*
964184	I	A->G	Between SACOL0961 *gluD* and SACOL0962 glycerophosphoryl diester phosphodiesterase
964192	I	A->G	Between SACOL0961 *gluD* and SACOL0962 glycerophosphoryl diester phosphodiesterase
1018644	NS	C->A	SACOL1011 hypothetical protein
1024540	I	C->A	Between SACOL1016 *fabI* and SACOL1017 hypothetical protein
1097696	I	G->T	Between SACOL1806 hypothetical protein and

			SACOL1087 hypothetical protein
1097788	I	G->A	Between SACOL1806 hypothetical protein and SACOL1087 hypothetical protein
1097822	I	T->G	Between SACOL1806 hypothetical protein and SACOL1087 hypothetical protein
1178314	NS	T->A	SACOL1170 hypothetical protein
1248844	I	C->T	SACOL1240 DAK2 domain protein
1300830	NS	G->A	SACOL1288 *infB*
1300880	NS	C->T	SACOL1288 *infB*
1300897	NS	G->A	SACOL1288 *infB*
1300903	NS	G->A	SACOL1288 *infB*
1300909	NS	A->G	SACOL1288 *infB*
1301378	NS	C->A	SACOL1288 *infB*
1546565	NS	T->A	SACOL1505 *aroB*
1589921	S	T->G	SACOL1553 glyoxalase family protein
1736839	S	T->C	SACOL1707 *radC*
1850520	I	G->C	Between SACOL1801 peptidase and SACOL1802 hypothetical protein
1977016	I	G->T	Between SACOL_tRNA-Pseudo-1 and SACOL_Sa23SD *rrlD*
2008915	I	G->A	Between SACOL1946 methionine aminopeptidase and SACOL1947 hypothetical protein
2019720	I	T->A	Between SACOL1958 hypothetical protein and SACOL1959 hypothetical protein
2019730	I	G->C	Between SACOL1958 hypothetical protein and SACOL1959 hypothetical protein
2019735	I	G->C	Between SACOL1958 hypothetical protein and SACOL1959 hypothetical protein
2074067	NS	T->C	SACOL2013 hypothetical protein
2079741	S	T->C	SACOL2019 *sdrH*
2082429	I	G->T	Between SACOL2021 hydrolase and SACOL2022 *hld*
2147052	NS	A->T	SACOL2081 hypothetical protein
2151680	NS	C->T	SACOL2087 hypothetical protein
2388474	I	G->T	Between SACOL2325 LysR family transcriptional regulator and SACOL2326 *fosB*
2388476	I	T->G	Between SACOL2325 LysR family transcriptional regulator and SACOL2326 *fosB*
2388477	I	C->T	Between SACOL2325 LysR family transcriptional regulator and SACOL2326 *fosB*
2388497	I	G->T	Between SACOL2325 LysR family transcriptional regulator and SACOL2326 *fosB*
2388617	I	T->G	Between SACOL2325 LysR family transcriptional regulator and SACOL2326 *fosB*
2432895	NS	A->T	SACOL2374 transcriptional regulator, TetR family
2433036	NS	A->G	SACOL2374 transcriptional regulator, TetR family
2441240	I	G->C	Between SACOL2381 hypothetical protein and SACOL2382 proton/sodium-glutamate symport protein
2561335	S	C->T	SA2505 *sasG*
2777260	I	C->T	Between SACOL2704 hypothetical protein and SACOL2705 hypothetical protein
2802679	I	A->T	Between SACOL2733 hypothetical protein and SACOL2734 hypothetical protein

3.6.3 Single acquisition of an SCC*mec* element leading to the emergence of MRSA

Analysis of the SCC*mec* variants indicated that they were closely related. In an attempt to elucidate if the SCC*mec* was acquired once by the archaic MRSA population, or multiple times, the phylogenetic relationship of the SCC*mec* and genetic background of the isolates used in the temporal analysis were compared.

Of the 122 isolates used for the linear regression analysis, two of the isolates had lost the *mecA* gene, including the Jevons isolate previously reported as methicillin sensitive (Table 3-2: RH10000388; Jevons ID 14083). An alignment of the 120 isolates was produced excluding the predicted regions of recombination in SCC*mec* (*pls*; intergenic between SACOL0030/31), as well as the single base site in IS*341* aforementioned. The subsequent 14 remaining SNP sites (Table 3-4) were then used to construct the phylogeny. Ten distinct haplotypes of the element were evident in the subpopulation derived across the eight-year period of 1960 to 1968. The relative abundance of each genotype as well as the genetic variation differentiating them is demonstrated as a minimum spanning tree in Figure 3-7(A). This revealed that the majority of the isolates could be assigned to a single haplotype (n=86), which included the Jevons isolates as the earliest MRSAs. Nine other related haplotypes of the type I element were identified suggesting that the majority haplotype may represent a founder population.

Table 3-4 Canonical SNPs in sub-group of 120 SCC*mec* type I elements, derived between 1960 and 1968. Base positions are relative to SCC*mec* in *S. aureus* COL reference (accession number: CP000046). NS: non-synonymous, S: synonymous, I: intergenic.

Base position	SNP type	Base change	Region/ gene
37775	NS	A->C	SACOL0030- conserved hypothetical protein
41826	S	C->T	SACOL0034- *mecR1*
42773	NS	G->A	SACOL0035- hypothetical protein
43712	NS	G->A	SACOL0036- IS*1272* transposase
47867	I	C->T	SACOL0041- *ccrB* frameshift
53309	S	G->A	SACOL0045- hypothetical protein
55179	Ns	C->T	SACOL0047- conserved hypothetical protein
55554	NS	C->T	SACOL0047- conserved hypothetical protein
62136	NS	A->G	SACOL0051- conserved hypothetical protein
62444	NS	T->G	SACOL0051- conserved hypothetical protein
63724	I	A->C	Between SACOL0051 and SACOL0052
65000	S	T->C	SACOL0052- glycosyl transferase, group 1 family protein
66076	I	G->A	Between SACOL0052 and SACOL0054 (pseudogene)
67458	I	T->A	Between SACOL0054 and SACOL0055

The distribution of these 10 SCC*mec* type I haplotypes throughout the population was then assessed by comparison with the core genome phylogeny (Figure 3-7(B)). Parsimonious reconstruction of the SCC*mec* haplotypes back on to the phylogeny, demonstrated that there was concordance between the core genome phylogeny and the SCCmec minimum spanning tree. This result suggests that the SCC*mec* element was introduced as a single horizontal transmission event corresponding with the basal node of the tree (Figure 3-7 (B)). The observed diversity in the MGE is therefore predicted to have arisen after its acquisition and subsequent association with the ST250/ST247 lineage. As demonstrated by the temporal analysis (Figure 3-6), the age of the SCC*mec* type I element in the archaic population can be traced back to 1946.

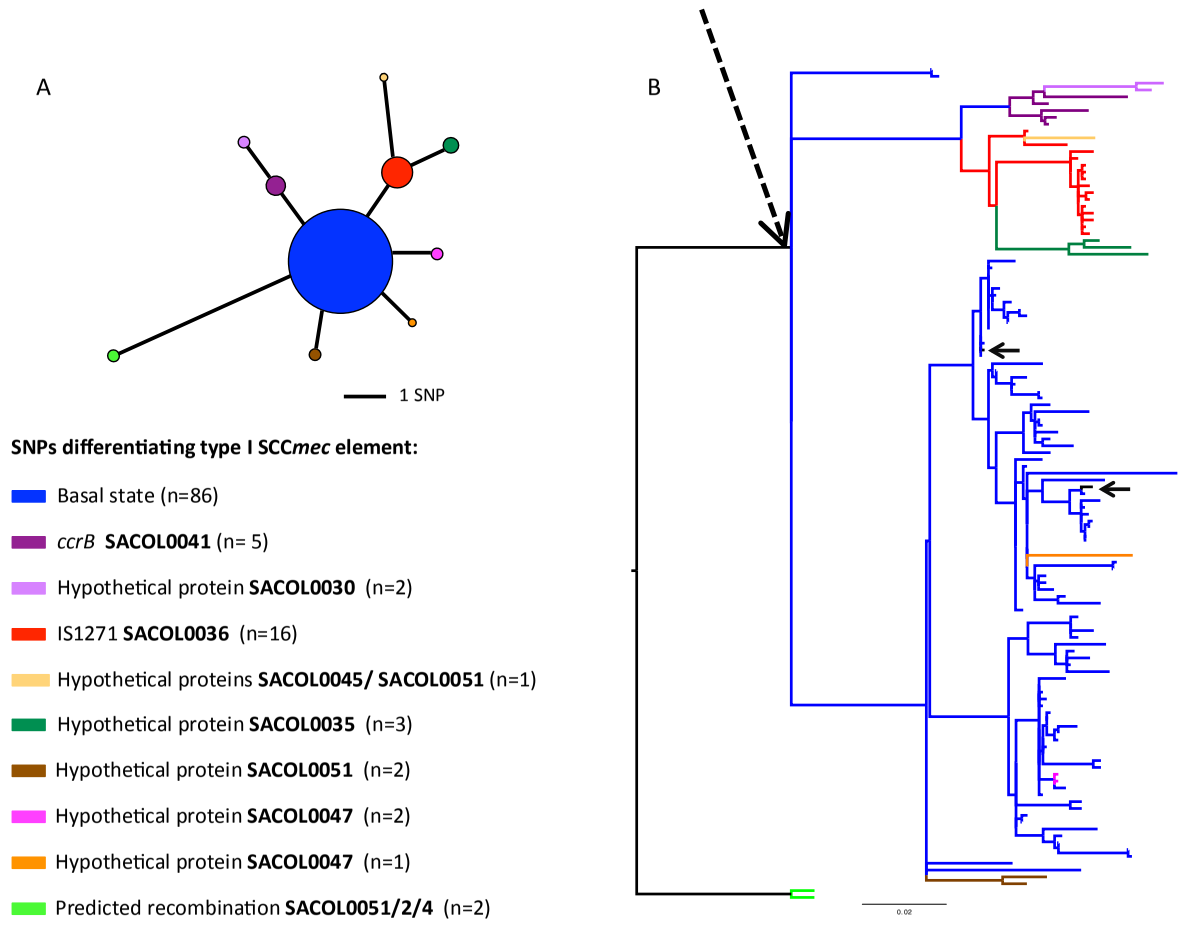

Figure 3-7 SCC*mec* type I haplotypes in the historic collection and their distribution within the population. A: minimum spanning tree of 120 type I elements from isolates used for temporal analysis. Tree was built using 14 SNPs across 120 SCC*mec* elements following removal of regions of recombination and IS*431mec* SNP. Relative volume of coloured circles correlated to number of isolates with SCC*mec* of that genotype. Colours represent the canonical SNP differentiating them from the majority state (in blue). **B:** Maximum likelihood core genome SNPs tree of 122 isolates used for temporal analysis. Branch colours correspond to the 10 SCC*mec* haplotypes carried by each isolate in the tree. Branch labels in black indicate two isolates that have lost SCC*mec* (also highlighted by black arrows). The perforated black arrow on basal branch indicates inferred entry point of SCC*mec* type I into the ST250/ ST247 population.

3.7 Discussion

This isolate collection has provided an unprecedented perspective on the evolution of the first known methicillin resistant *S. aureus* lineage. In this collection, preserved for decades by the institutions who first detected them were the very first identified representatives of MRSA. Genomic analysis of these isolates has allowed the evolution of the archaic clone of MRSA to be retraced, and has allowed the questions of when SCC*mec* first entered *S. aureus*, as well as on how many occasions, to be addressed. It has additionally allowed a retrospective view of how antimicrobial pressures applied at the time shaped this historical lineage, with important lessons being highlighted in the context of the current challenges of antimicrobial resistance.

The evolutionary origins of MRSA have previously been traced to *S. aureus* CC8 (Enright et al. 2002). The descendent of the ST8 MSSA, ST250 became the first strain to acquire SCC*mec*. This study also demonstrated that ST247 MRSA originated from ST250 though a single point mutation gained in the *gmk* allele. Evident from the core phylogeny in Figure 3-1, the ST247 population is seen to emanate from the basal ST250 isolates in keeping with the hypothesis that ST247 MRSA evolved from an ancestral ST250-MRSA-I. The initial ST250 strain has been proposed to have been an MSSA before its acquisition of the SCC*mec* (Enright et al. 2002). The *mecA* negative strains identified in the collection are represented throughout the phylogeny (Figure 3-3), noted in both ST250/ST247 backgrounds. This indicates they are therefore not representatives of the inferred ancestral MSSA state of this lineage, but are instead isolates that have become so through loss of the SCC*mec*.

Although not decisively concluded, the origins of SCC*mec* are almost certainly from one of the coagulase negative staphylococci (Couto et al. 1996; Wisplinghoff et al. 2003; Barbier et al. 2010). In *S. aureus* the ST250 strain background appears to have been the first recipient of the element. The issue of whether this MGE entered as an isolated event or on multiple occasions

has also never been resolved. The penicillin resistance determinant *mecA* has been hypothesised to have entered *S. aureus* as a solitary event with its eventual propagation throughout the species (Kreiswirth et al. 1993). Analysis of the variation within the SCC*mec* of this population was specifically undertaken to clarify this. The observed diversity was predominantly within the region of the *pls* gene (Figure 3-4), which has been formerly been reported for its variability within the type I element (Shore et al. 2005). The function of this 230 kDa cell wall anchored (CWA) protein is not fully understood, but has been shown to promote adherence to desquamated nasal keratinocytes as well as being associated with reduced cellular invasiveness (Roche 2003; Juuti et al. 2004). It is also interestingly noted for its presence in methicillin resistant coagulase negative species (Juuti et al. 2005). The highly repetitive nature of this LPxTG protein, with serine/ aspartate rich structure, makes it a region susceptible to homologous recombination. As known from other lineages, CWA proteins are subject to the effects of diversifying selection and are frequently mutated (Basic-Hammer et al. 2010; Holden et al. 2013). This is as also evidenced by the regions of recombination identified within the core genome of the collection in which numerous CWAs such as the clumping factors (*clfA*/*B*) were mutated (Table 3-1). Exclusion of variation due from recombination in SCC*mec* during construction of the phylogeny demonstrated that 116 of the isolates were carrying identical elements (Figure 3-5), whilst relatively short genetic distances (maximum of 8 SNPs) separated the remainder (n=76) given their temporal divergence of up to 29 years apart. Subsequent comparison of the distribution of the SCC*mec* variants with the core genome phylogeny (Figure 3-7(B)) revealed congruence, indicating that the SCC*mec* type I element was acquired once by the ST250 background.

In the first description of the type I SCC*mec* element it was noted that there was a frameshift mutation in *ccrB1* recombinase (Ito et al. 2001). The *ccr* recombinases are required for both integration and excision of the element from the chromosome (Katayama et al. 2000), with *ccrB* being specifically required for excision (Wang and Archer 2010). The mutation identified

in the NCTC10442 type I SCC*mec* is predicted to result in a non-functioning recombinase (Noto and Archer 2006). All of the type I elements in this study have the NCTC10442-type *ccrB* frameshift, which would give further credence to the hypothesis that SCC*mec* was introduced into this population once. The initial integration event was then presumably followed by the development of this frameshift mutation, which consequently led to fixation in the lineage due to dysfunctional recombinase apparatus.

A single instance of an alternative type of SCC*mec* in the collection was identified. This was notably the TLV used as the out-group for the analysis (ID: RH12000692_7401696; online supplementary data file Chapter 3-Historic MRSA Suppl. data)(Figure 3-1). Comparison of this MGE to currently described SCC*mec* elements revealed it to be a type IVh SCC*mec* (Milheiriço et al. 2007). First described in 2002 this SCC*mec* type is now synonymous with the endemic MRSA clone EMRSA-15 (Ma et al. 2002; Holden et al. 2013). The smaller size and alternate recombinase apparatus has seen the SCC*mec* type IV elements become widely distributed including in the highly successful EMRSA-15. The classification of this element is in part based upon the order they were first observed. Type IV elements have been identified in the coagulase negative species *Staphylococcus epidermidis*, and in isolates dating from the 1970s (Wisplinghoff et al. 2003). The identification of a type IV element in this collection is of interest, as it would pre-date the previous earliest reports of this element type in *S. epidermidis* (Wisplinghoff et al. 2003). It has been suggested that in *S. aureus* the earliest presence of the type IV element was within ST5 Paediatric clone in North America during the late 1980s and 1990s (Lina et al. 2006; Chambers and DeLeo 2009a). Additionally this would be in keeping with the type IV elements proposed evolutionary origin from the type I SCC*mec* (Lina et al. 2006).

Estimating the time point at which the archetypal MRSA lineage arose was one of the primary aims of the study. The earliest reports of methicillin resistance came from the UK and Denmark

within a year of its first use in both countries respectively (Jevons 1961; Eriksen and Erichsen 1964; Jessen et al. 1969). It has therefore always reasonably been concluded that resistance emerged after the drugs introduction and as an adaptive measure following exposure to it. The temporal analysis applied to the strain background (Figure 3-6) dates the time to emergence of the common ancestor of this population at 1946, fourteen years before the first use of the second-generation beta-lactam. As a more robust assessment of the date of emergence of this background using BEAST (Drummond et al. 2012) supports the most recent common ancestor as being 1946 (95% highest posterior density (HPD) 1938-1952; M.T.G. Holden, personal communication). In a historical context, these dates fall within a period in history when penicillin resistance was becoming a concern (Kirby 1944), only 4 years after its introduction having transformed the treatment of staphylococcal infection (Chain et al. 1940).

Evident from the resistance profiles of the isolates, penicillin resistance was very widely distributed throughout the population in this period, with 86% of isolates carrying the beta-lactamase *blaZ*. Carriage of *mecA* confers the recipient with an alternate mechanism by which to resist penicillin, via the auxiliary penicillin binding protein PBP2a. This determinant was identified more frequently than *blaZ*, with 94% of the isolates carrying it. A total of 180 (86%) of the isolates are noted to carry both *mecA* and *blaZ*. There is presumably additional advantage bestowed by carriage both penicillin resistance mechanisms, despite one being on a sizeable element (SCC*mec* type I - 34 kb) with no other associated resistance genes. At the time of the first recognition of MRSA in 1960, PHE had found only 3 resistant strains from more than 5000 screened (Jevons 1961). This would suggest overall low population incidence of MRSA at the time. It therefore appears that when methicillin was introduced to circumvent penicillin resistance, it didn't lead to the emergence of MRSA, but in fact selected for the as yet unrecognised methicillin resistant *S. aureus.* Thus the novel drug became the selective pressure driving the emergence of MRSA.

The extent of selective pressure exerted by antimicrobial use in this era on this *S. aureus* clone is evident from the collection. For instance tetracycline resistance is very widely distributed throughout the isolates, *tetK* is present in 89.4% and *tetM* in a further 44.9%. This prevalence clearly reflects historical prescribing practice, with the first of this class of antibiotic trialled clinically in 1948 (Duggar 1948). Tetracycline resistance is now considered a feature of livestock associated MRSA clones (Price et al. 2012), with its overall incidence being uncommon otherwise (Aanensen et al. 2016). Notable by their relative paucity however are acquired resistance determinants to Streptomycin. This resistance is another defining feature of the ST250/247-MRSA-I lineage. Two acquired streptomycin resistance markers were identified in the population, *str* and *aad9*. Only a single instance of *str* was found in the collection, whilst *aad9* was present in 39.7% of isolates. Both of these determinants are noticeably absent form the Jevons isolates, which characterise this lineage.

In Mycobacterium tuberculosis streptomycin resistance has been shown to be caused by a point mutation in the *rpsL* gene (lysine43arginine), which encodes ribosomal protein S12 (Zhang and Young 1994). Subsequent comparison of the equivalent residue in the *rpsL* gene of the ST250/ ST247 isolates (n=208) revealed an arginine residue at position (codon) 56. The exception being the TLV outlier (ID: RH12000692_7401696), which had a lysine at this position (M.T.G. Holden, personal communication). Comparison of this residue in other *S. aureus* reference genomes such as MRSA252 and MSSA476 also revealed a lysine in this position, suggesting that the lysine residue represents the primordial residue. This mutation is therefore likely to account for the reported streptomycin resistance in this lineage (Crisóstomo et al. 2001). This would be supported by a recent technical application of *rpsL* mutations in *S. aureus* to allow selection of mutants during allelic exchange (Chen et al. 2015).

Outwith the previously noted core resistance profile of this lineage (Penicillin/Methicillin/Streptomycin/Tetracycline) determinants of resistance to a further 6

classes of antibiotics are present. These include chloramphenicol, macrolides (erythromycin), aminoglycosides (kanamycin), aminocoumarins (novobiocin), sulphonamides (trimethoprim), and fusidic acid. Some are more widely distributed than others, for instance *ermA* (erythromycin resistance), which is notably present in each instance of *AAD9* (spectinomycin resistance). This reflecting its co-carriage on the MGE Tn*554* (Phillips and Novick 1979; Murphy et al. 1985). The kanamycin resistant determinants (*aphA-3/aadD*) are only found in isolates dating from 1967 onwards and are not found in any of the Danish isolates. Whether this reflects drug use at the time outwith the UK it is not possible to conclude, but is interesting as kanamycin resistance in *S. aureus* was noted for its increasing incidence as early as 1958 in the USA (Griffith 1966) in the context of the drug's reported discovery in 1957 (Umezawa et al. 1957).

The point mutations found in the collection are associated with the resistance genotypes to novobiocin, fusidic acid and trimethoprim. Identified as one of the homoplastic SNPs (Table 3-3) in the collection the mutation in DNA gyrase B (*gyrB,* R144I) has been reported to provide resistance to the now long retired antibiotic novobiocin (Stieger et al. 1996; Bisacchi and Manchester 2014). The remaining two point mutations identified are associated with resistance to trimethoprim (dihydrofolate reductase *dfrA,* F99Y) and fusidic acid (translation elongation factor G *fusA,* P404L) in a single instance each. These also incidentally represent the last antibiotics to be introduced in the context of the resistances seen in this collection, having been first used in 1960 and 1962 correspondingly in the UK (Newman et al. 1962; Noall et al. 1962).

3.8 Concluding remarks

This analysis reveals that *S. aureus* acquired SCC*mec* long before the introduction of methicillin, and as a single horizontal event. Introduced into a penicillin susceptible

background the recipient was provided with the selective advantage of being able to circumvent the increasingly used β- lactam antibiotic. The emergence of MRSA is seemingly the unforeseen consequence of the extensive use of penicillin in the 1940s. This highlights that the potential for new antimicrobial agents to be rendered ineffective because of prior adaptations in the organism in response to previous selective pressure. This in itself remains one of the major challenges in tackling antimicrobial resistance.

4 *Staphylococcus aureus* carriage in healthy children

4.1 Introduction

Whole genome sequencing (WGS) has transformed our understanding of *S. aureus* as a pathogen and allows the evolution of the organism's genome during carriage in a host to be explored. The genomic diversification occurring with the host has been demonstrated by multiple studies, many of which have identified subsequent consequences for the disease causing potential of the organism (Young et al. 2012; Harris et al. 2013; Paterson et al. 2015). Examination of the within-host diversity allows us to investigate how populations arise, differentiate and adapt in a host, and to reconstruct the evolutionary events that shape a colonizing population.

The DNA library preparation and whole genome sequencing for this study was kindly carried out by Dr Kerry Pettigrew, Infection Group, University of St Andrews.

4.2 Aims and objectives

The aim of this work was to characterise *S. aureus* strains colonising healthy children in the community. Whilst diversity has been assessed in the adult population, and predominantly in healthcare associated settings relating to transmission of hospital adapted MRSA clones, this work aimed to define in host diversity within asymptomatically colonised children. Each individual's colonising population was examined to define this fine-scale genetic variation specifically looking for evidence of adaptive mutations and their potential biological consequences. The characterisation of these community controls was then intended for direct comparison with colonisation of children affected by the inflammatory disorder atopic eczema presented in Chapter 5.

4.3 Recruitment, participation and colonisation rates

Recruitment was through local schools and a single nursery that were within the Tayside and North Fife area. This was to allow for a control population within the same geographical catchment area as AE cases that were recruited through Ninewells Hospital. Multiple study sites were included to allow sufficient and age-matched recruitment. This was also with the intention of obtaining samples from unlinked individuals. Sampling from siblings, and those in close contact in the school environment could subsequently influence carriage strain and observed diversity in the partcipants.

Children between the ages of 0-12 years were invited to participate. Participants parents were asked to complete a study questionnaire providing information on factors which may influence both carriage of *S. aureus* and strain type carried including: past medical history, recent antimicrobial therapy, hospital attendance (inpatient or outpatient) within a year. During recruitment visits to each site children with written parental consent and who verbally assented to participation were recruited. An examination of the skin of the forearms and popliteal fossae was undertaken to look for evidence of inflammatory skin disease that may also influence *S. aureus* carriage. Each participant was then swabbed from a single nostril, and from a single antecubital fossa to assess for both nasal and extra-nasal colonisation.

Summarised in Figure 4-1 is the study recruitment process to gain 306 child participants across 4 schools and 1 nursery in the Tayside and North Fife area. Carriage rates were variable across the study sites and were as follows: school 1 (29.1%), school 2 (34.0%), school 3 (30.9%), school 4 (42.8%), and the single nursery site (8.3%). The highest carriage rate was noted in school 4 which also notably had the largest student group (n=500). This carriage rate may reflect the density of the student population at this study site as crowding is a known risk factor for colonisation, but rates of between 40 and 50% have been reported in children

between 6 and 12 years old (Sollid et al. 2014). The low carriage rate noted at the nursery site was potentially influenced by high rates of the study participants being affected with the common cold at the time of sampling (January/ February 2016), which made swabbing more technically difficult in this younger population. However *S. aureus* carriage rates in children of 14 months have been reported at rates of 12.9% (Lebon et al. 2008).

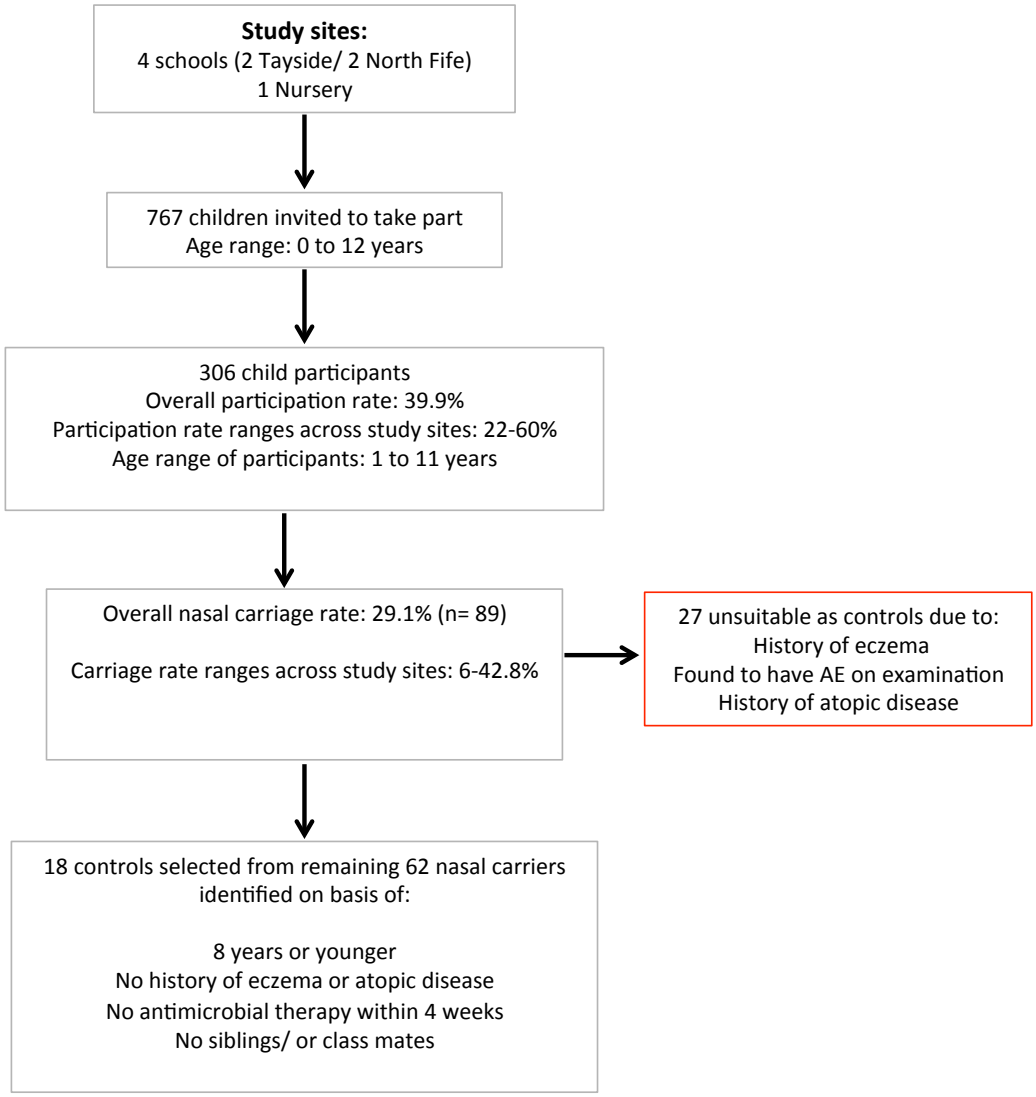

Figure 4-1 Community control study recruitment flow chart. Figures in brackets indicate ranges across five recruitment study sites for both participation and *S. aureus* nasal carriage.

The overall nasal carriage rate detected was 29.1% in recruits between the ages of 1 and 11 year of which 46% (n= 41) were female and 54% (n= 48) were male. No instances of extra-nasal colonisation were identified outwith the context of active inflammatory skin disease in the sampled controls. It should be acknowledged that this study did not attempt to sample from the other common extra-nasal carriage sites such as groin or perineum (Wertheim et al. 2005). This was intentional in the study design, as it was felt unlikely that school children or parents would consider sampling from these sites within this study setting.

4.3.1 Sampling method and rationale

From nasal samples obtained, 18 were selected as controls for assessment of the diversity in their colonising population. This number was to allow comparison of twice as many controls as cases (presented in Chapter 5). This decision was also taken on the basis of feasibility of cost relating to WGS of the samples. Multiple colonies were selected per control for sequencing to allow assessment of in host variation. From each control a total of 5 colonies were sequenced given previous evidence that 3 or more colonies must be assessed to accurately identify of co-colonisation (Votintseva et al. 2014).

4.4 Clinical phenotype of recruits

Eighteen controls were selected from the 89 nasal carriers identified across 4 school study sites. Isolates from recruits with a history of atopic disease (inclusive of atopic eczema, hayfever, asthma and food allergy), or recent antimicrobial therapy were not utilised for comparators to the AE cases. The features of the selected controls are shown in Table 4-1.

Table 4-1 Nasal colonisation control characteristics.

Study ID	Age	Sex	School site
SS_016	5	F	1
SS_026	6	M	1
SS_039	7	M	1
SS_045	7	F	1
SS_057	8	M	2
SS_059	8	F	2
SS_091	5	F	2
SS_094	6	M	2
SS_099	7	F	2
SS_105	7	M	2
SS_141	5	F	3
SS_147	6	M	3
SS_149	6	F	3
SS_153	7	M	3
SS_157	7	F	3
SS_250	8	F	4
SS_268	8	M	4
SS_303	6	M	4

4.5 Isolation of S. aureus from clinical samples

4.5.1 Microbiological burden

The recovered bacterial burden in the 18 healthy children selected as controls from the larger carriage (n=89) population is presented in Table 4-2. Carriage burden was clearly variable, ranging from light growth in participants such as SS_091 and SS_153 to heavy growth in participants such as SS_045 and SS_039. The recovered CFU counts and corresponding colony counts for the 18 selected controls are shown in Table 4-2. These are representative of colonies derived before selective enrichment, with the exception being SS_303 whose initial sample was negative prior to 18 hours enrichment growth in nutrient broth.

Table 4-2 Bacterial burden in nasal carriage controls. CFU counts from 18 controls. (*colonies were derived from enrichment).

Study ID	CFU/ml
SS_016	1.28×10^{6}
SS_026	1.03×10^{6}
SS_039	4.0×10^{6}
SS_045	$> 5.0 \times 10^{6}$
SS_057	4.7×10^{5}
SS_059	1.5×10^{5}
SS_091	2.0×10^{4}
SS_094	3.0×10^{4}
SS_099	9.1×10^{5}
SS_105	3.87×10^{6}
SS_141	4.4×10^{5}
SS_147	3.7×10^{5}
SS_149	8.9×10^{5}
SS_153	2.0×10^{4}
SS_157	3.3×10^{5}
SS_250	3.0×10^{4}
SS_268	3.1×10^{5}
SS_303	$> 5.0 \times 10^{6}$ *

4.5.2 Confirmation of *S. aureus* from clinical samples

Presumptive *S. aureus* colonies obtained by selective media were then sub cultured onto BHI agar and then confirmed by colony PCR for detection of the presence of *femB* gene (Figure 4-2). Confirmed *S. aureus* colonies were then grown overnight for DNA extraction and stocking at -80°C.

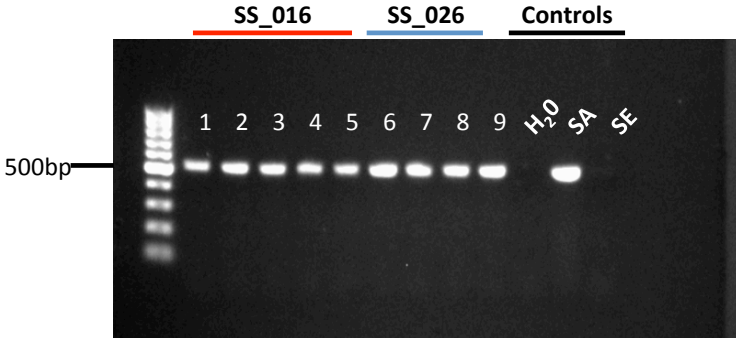

Figure 4-2 *femB* Colony PCR confirming *S. aureus* from clinical samples. Nasal colonies numbered between 1 -5 from study recruit SS_016 and 6-9 in recruit SS_026. Controls: H_20, genomic DNA from *S. aureus* (SA) and *S. epidermidis* (SE). 500bp indicates size of product.

4.6 Genetic background of colonising strains

A total of 90 single colony isolates from 18 study participants (5 per individual) were then sequenced. From the assembled genomes the MLST profile was extracted revealing that colonisation was by a single strain type in each individual. No instances of co-colonisation were found despite multiple colonies being assessed. As shown in Table 4-3 isolates belonging to clonal complex 30 were the most frequently identified, found in 11 of 18 (61.1%) of the controls. The remaining controls were carrying strains belonging to CC15 (n=3, 16.7%), CC1 (n=1; 5.5%), CC5 (n=1; 5.5%), CC45 (n=1; 5.5)% and CC22 (n=1; 5.5%).

Table 4-3 Colonising strains in nasal carriage controls. Sequence type (ST), clonal complex (CC), and MLST (multi-locus sequence type) profile is representative of all 5 sequenced colonies for the individual. (*arc*- carbamate kinase 1; *aroE*- shikimate dehydrogenase; *glpF*- glycerol kinase; *gmk*- guanylate kinase; *pta*- phosphate acetyltransferase; *tpi*- triosephosphate isomerase; *yqil*- acetyl conenzyme A acetyltransferase

Study ID	ST	CC	*arcC*	*aroE*	*glpF*	*gmk*	*pta*	*tpi*	*yqil*
SS_016	45	45	10	14	8	6	10	3	2
SS_026	30	30	2	2	2	2	6	3	2
SS_039	30	30	2	2	2	2	6	3	2
SS_045	15	15	13	13	1	1	12	11	13
SS_057	582	15	13	13	1	1	12	10	13
SS_059	30	30	2	2	2	2	6	3	2
SS_091	30	30	2	2	2	2	6	3	2
SS_094	5	5	1	4	1	4	12	1	10
SS_099	30	30	2	2	2	2	6	3	2
SS_105	39	30	2	2	2	2	2	2	2
SS_141	22	22	7	6	1	5	8	8	6
SS_147	30	30	2	2	2	2	6	3	2
SS_149	30	30	2	2	2	2	6	3	2
SS_153	582	15	13	13	1	1	12	10	13
SS_157	30	30	2	2	2	2	6	3	2
SS_250	30	30	2	2	2	2	6	3	2
SS_268	2889	30	324	2	2	2	6	3	2
SS_303	1	1	1	1	1	1	1	1	1

4.6.1 ST30 comparison

One of the considerations when selecting control samples was that cross-transmission between individuals within study sites may influence the observed diversity as has been demonstrated by other studies characterising in host diversity and transmission (Harris et al. 2013; Tong et al. 2015). Controls were selected to avoid the inclusion of individuals who would be in close contact such as siblings and classmates. Evident from the strain types recovered in addition to the abundance of CC30 derived strains, there were individuals with ST30 isolates recruited from the same school site (Table 4-3). We wished to examine whether the observed abundance of ST30, as well as any potential in host diversity found had been influenced by cross transmission events within classes or schools. ST30 isolates from the 3 school sites with multiple recruits were used to construct a phylogeny with comparison of the SNP distances between the colonising populations of different individuals.

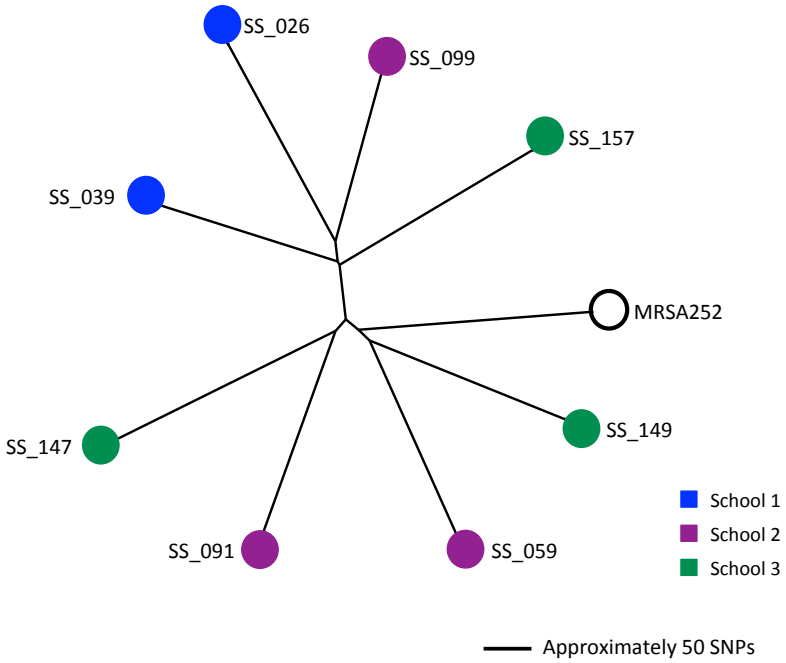

Figure 4-3 ST30 isolate comparison across recruitment sites. Unrooted maximum likelihood core SNPs tree comparing the relative genetic distances of 5 single colony isolates from 8 controls across 3 schools. Colours on branches correlate to the school attended by the control, as indicated by legend. Labels next to branches indicate the study ID of the control. *S. aureus* MRSA252 used as reference chromosome for construction of phylogeny. Scale bar indicates approximate genetic distance.

As show above in Figure 4-3 ST30 isolates from 8 participants across 3 school sites within a 10-mile radius were compared. Obvious from the phylogeny is that isolates both within and across schools exhibited significant diversity. Isolates from participants in a single school are seen to differentiate from one another by similar genetic distances to those derived from other geographically distinct school sites. The shortest SNP distance separating any of the individual controls isolates within any one of the school sites was 385 core SNPs, in the instance of SS_099 and SS_059. The greatest distance noted within samples from a single study site was 506 core SNPs (between SS_147 and SS_157). This is similar to the greatest distance between any two single ST30 colonies between the 8 controls of 513 core SNPs (SS_039 and SS_147), which were from North Fife and Dundee respectively. All isolates differentiated from the ST36 reference by at least 450 core SNPs, which demonstrates that

these ST30 isolates are genetically diverse, originating from a well-established community population. Previous reports of in host diversity and transmission events have shown that SNP distances of up to a maximum of between 40 and 60 SNPs can be used to distinguish related isolates (Young et al. 2012; Tong et al. 2015). The genetic distances separating these controls isolates would therefore suggest that the carriage of ST30 and the observed diversity was not the result of cross-transmission between recruits at the same site.

4.7 Resistome profile of the colonising populations

Another potential source of genetic variation in the *S. aureus* genome are mobile genetic elements (MGEs) conferring antimicrobial resistance. Variability in the carriage of such elements has previously been shown to occur during asymptomatic carriage (Harris et al. 2013; Stanczak-Mrozek et al. 2015) The resistance profile of all 90 colonies from the 18 controls was assessed to determine if any such variability had arisen.

The sequence reads were mapped against a reference allele of known resistance determinants (Table 2-15; Materials and Methods) using SRST2. The results of this demonstrated that there were no instances of variability in resistance determinants carried in the colonising population of individuals. Table 4-4 shows the resistance profiles for each of the 18 controls. Carriage of penicillin resistance determinant *blaZ* was universal, whilst other clinically relevant determinants were present more sporadically. Single instances of resistance determinants to fusidic acid (*fusC*), erythromycin (*ermA*), chloramphenicol (*fexA*) and spectinomycin (*ADD9*) were found. Associated with reduced susceptibility to disinfectants, *qacC*, was found in 3 controls. Heavy metal resistance determinants were also evident throughout the population. The ST30 and ST22 isolates were all found to be carrying the arsenical resistance genes (*arsB/C*) and copper resistance determinant *mco* which are known to be co-carried on an integrated plasmid in MRSA252 (Holden et al. 2004) or SAP019A (accession number

GQ900385.1). Cadmium resistance determinant *cadD* was present in multiple controls, and is also know to be co-carried on plasmids such as pSAS1 in MSSA476 (Holden et al. 2004).

Table 4-4 Resistance profiles of community carriage control isolates. Results per individual are representative of all 5 colonies sequenced. No in-host variation on AMR genes was identified. Boxes indicated with – denote absence of resistance gene. (Gene abbreviations are as defined in Materials and Methods section 2.6.9, Table 2-15).

Study ID	ST	*blaZ*	*fusC*	*fexA*	*qacC*	*AAD9*	*ermA*	*arsB*	*arsC*	*cadA*	*mco*	*cadD*
SS_016	45	*blaZ*	-	-	-	-	-	-	-	-	-	*cadD*
SS_026	30	*blaZ*	-	-	-	*AAD9*	*ermA*	*arsB*	*arsC*	*cadA*	*mco*	-
SS_039	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_045	15	*blaZ*	-	-	-	-	-	-	-	-	-	*cadD*
SS_057	582	*blaZ*	-	-	*qacC*	-	-	-	-	-	-	*cadD*
SS_059	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_091	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_094	5	*blaZ*	-	*fexA*	*qacC*	-	-	-	-	-	-	*cadD*
SS_099	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_105	39	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_141	22	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_147	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_149	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_153	582	*blaZ*	-	-	-	-	-	-	-	-	-	*cadD*
SS_157	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_250	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_268	2889	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*cadA*	*mco*	-
SS_303	1	*blaZ*	*fusC*	-	*qacC*	-	-	-	-	-	-	*cadD*

4.8 Genetic heterogeneity in S. aureus colonising healthy children

4.8.1 Diversity of colonisation

Having identified that each control was colonised by a single strain, we sought to assess for evidence of diversification arising during asymptomatic carriage. In order to capture the genetic diversity each of the control's samples were mapped to a reference chromosome sequence CC where available, or draft assembly from one of the colonies where a reference genome was not available. This allowed identification of heterogeneity in the form of SNPs within each individual's 5 sequenced colonies to be determined. To quantify the relative within host diversity across this control population the number of core genome SNPs per sequenced colony was calculated. This demonstrated that diversities within individuals ranged from zero

SNPs per colony through to a maximum of 3.4 SNPs per colony in the most diverse control (SS_149) (Figure 4-4).

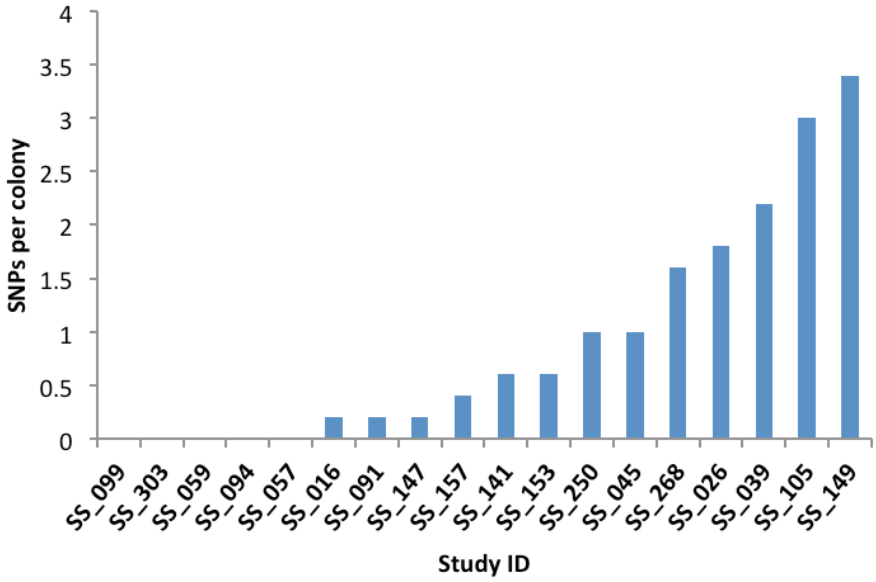

Figure 4-4 In-host diversity levels in nasal carriage controls. Diversity calculated as number of core genome SNPs/ sequenced colony/control.

4.8.2 Estimating the age of the colonising population

In order to estimate the age of the colonising population the maximum pairwise distance between any two colonies from each individuals colonising population was calculated. This represents the largest genetic difference in the population sampled, and was used to estimate the time period over which the most divergent colonisies sampled diverged from a common ancestor, utilizing a mutations rate median of 1.6 $x10^{-6}$ base substitutions/site/year that had previously been determined (Uhlemann et al. 2014a). The calculations were based upon substitution rates in the core genome, after accessory regions were removed. The size of the core genome for each reference chromosome used in this study is shown (Table 4-5).

Table 4-5 Size of core genome from representative reference genomes and self-assemblies used for mapping and SNP assessment. Estimated SNP rates for each lineage are based upon size of core genome. *No regions were excluded as accessory from self-assemblies and calculations were based on the total assembly size.

Reference genome/ assembly	Assembly size (bp)	Accessory genome regions excluded (bp)	Core genome size (bp)	Estimated SNPs/ genome/ month
MSSA476	2799802	172174	2627628	0.35
MRSA252	2902619	266437	2636182	0.35
HO 5096 0412	2832299	205466	2659077	0.35
CA347	2850503	223275	2627228	0.35
N315	2814816	205466	2609350	0.35
153 NC4*	2678675	N/A	2678675	0.36
045 NC1*	2699309	N/A	2699309	0.36

This suggested that the age of within individual populations sampled in the controls ranged from less than 1 week to a maximum of 24 months (Figure 4-5). Five controls were predicted to have been colonised by populations less than 1.4 months old. Eight were estimated to have been colonised by populations of between 1.4 and 7 months old. The remaining five had colonisation populations estimated to have arisen 10 to 24 months previously.

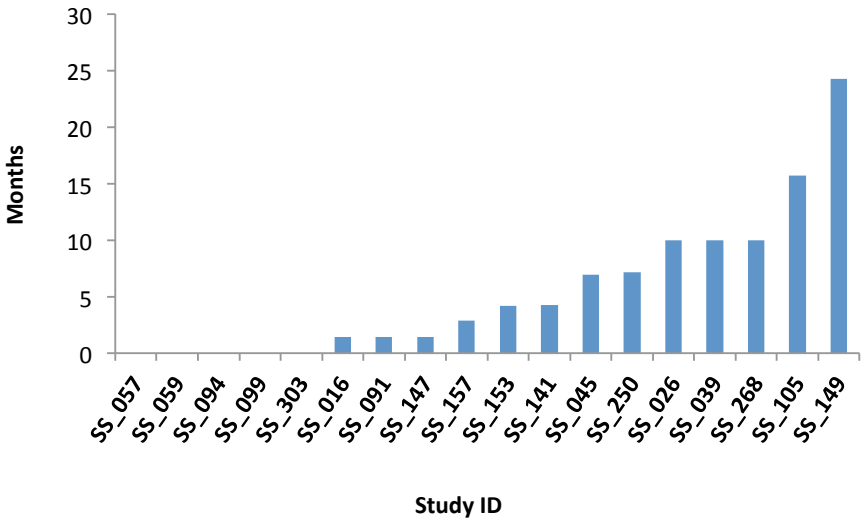

Figure 4-5 Estimated age of the colonising population in asymptomatic carriage controls. Time to most common recent ancestor of population is presented in months and is calculated from the maximum pairwise SNP distance observed in the sampled populations, and is estimated using a mutation rate of 1.6 x10^{-6} base substitutions/site/year, on basis of the core genome size presented in Table 4-5.

4.8.3 Diversity within individuals

The genetic diversity having been quantified across the control population, each controls samples were then assessed to specifically define the variation. To gain a high-resolution view of this observed diversity, the core genome SNPs were reconstructed back onto the phylogeny of each control allowing SNPs to be viewed in the context of their position within the phylogeny.

Overall 5 controls were colonised by genetically homogenous populations with all of their respective colonies being identical at core genome level (study IDs: SS_057/ 059/ 094/ 099 and 303) (Figure 4-4). In the remaining 13 controls, the number of SNPs differentiating the 5 colonies within any individual ranged from between 1 and 17 SNPs. These levels of diversity are in keeping with previously reported ranges described in the context of carriage and transmission (Golubchik et al. 2013; Tong et al. 2015).

The following results are representative of the three most diverse populations observed in the asymptomatic carriers. The phylogeny, core genome SNPs and indels identified in each of the remaining controls are included within the supplementary data appendix (Appendix A).

4.8.3.1 Control study participant SS_039

In the colonies sequenced from this 7-year-old boy the observed diversity had suggested the colonising population arose 10 months previously (Figure 4-5). A total of 11 SNPs distinguished the 5 colonies, with a maximum SNP distance between any two colonies of 7. As shown from the phylogeny (Figure 4-6) nose colonies 2 and 4 (NC2/4) are genetically identical separated from the remaining 3 colonies by a single SNP. The remaining three colonies exhibited more diversity, placing on longer branches and with 3 (NC1/5) or 4 (NC3) unique SNPs differentiating

them from the rest of the population. The SNPs identified between colonies are described in Table 4-6-1. A single deletion was found to differentiate the colonies, being absent in only NC5, and is shown in Table 4-6-2.

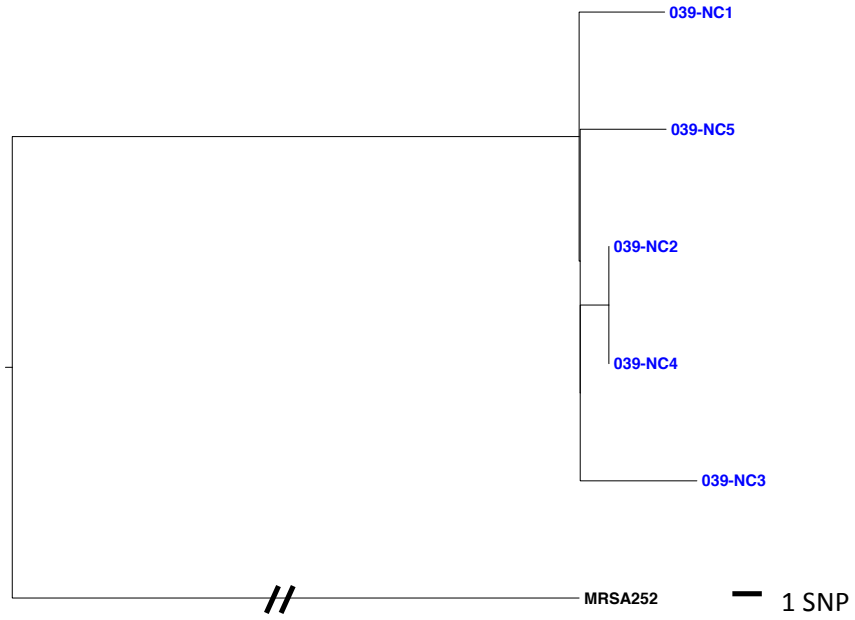

Figure 4-6 Maximum likelihood core SNPs tree for control study ID SS_039. Tree of 5 sequenced nasal colonies. Colonisation in this individual was by an ST30 strain. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table 4-6-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_039. Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position in ST30 reference (MRSA252; accession number: BX571856). NS– non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
2193102	G->A	NS	SAR2132 membrane protein	A->V	NC3
2584058	T->A	NS	SAR2509 gamma haemolysin A precursor	V->E	NC3
239541	A->G	NS	SAR0206 extracellular sugar-binding lipoprotein	D->G	NC2/4
2695006	A->G	NS	SAR2610 putative L-serine dehydratase, alpha chain	G->D	NC5
2737498	C->T	NS	SAR2649 membrane protein	V->I	NC5
31404	C->T	NS	SAR0023 S*asH*; CWA and nucleotidase	P->L	NC3
1668111	T->C	S	SAR1594 *bfmB1B* 2-oxoisovalerate dehydrogenase beta subunit		NC1
1033383	T->C	S	SAR0988 *murE* UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase		NC3
1775262	A->G	I	between SAR1712 putative N-acetylmuramoyl-L-alanine amidase and SAR1713 hypothetical protein		NC1
2872382	G->A	I	between SAR2768 hypothetical protein)and SAR2769 hypothetical protein		NC1
2717533	G->A	I	between SAR2631 hypothetical protein and SAR2632 transport protein		NC5

Table 4-6-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_039. Unique indels in the core genome differentiating sequenced colonies. Base position in relation to reference assembly used for mapping (039-NC5) and region/ gene are relative to position in ST30 reference genome (MRSA252; accession number: BX571856). (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
969680	NC1, 2,3,4	(-) att	intergenic between SAR2135 membrane protein and SAR2134 ABC transporter ATP-binding protein	65bp upstream of translational start site in SAR2135 - in putative promoter region

4.8.3.2 Control study participant SS_105

The analysis of this 7-year-old male's colonising population gave another example of an individual predicted to have been colonised for a prolonged period. In this case the age of the carriage population was estimated at 15.7 months (Figure 4-5). There were overall 15 core genome SNPs found between the 5 colonies, with the maximum SNP distance between any two colonies being 11. As shown in the phylogeny, all colonies were genetically distinct, being separated by at least a single SNP (Figure 4-7; Table 4-7-1). A total of 5 indels were found amongst the colonies, with 3 being unique to NC5, which is also noted to be on the longest branch length (Table 4-7-2).

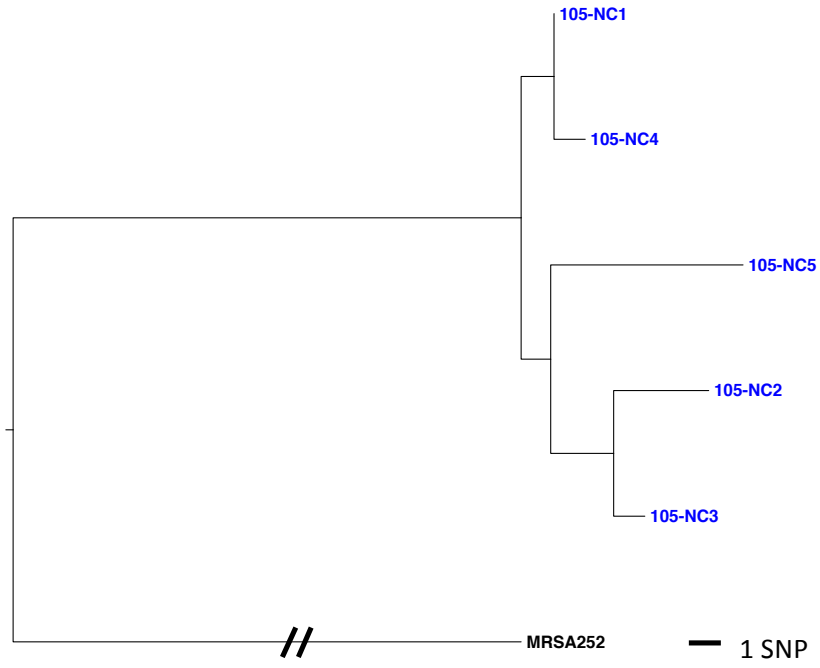

Figure 4-7 Maximum likelihood core SNPs tree for control study ID SS_105. Tree of 5 sequenced nasal colonies. Colonisation in this individual was by an ST39 strain. Tree rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table 4-7-1 Within host-heterogeneity identified in nasal carriage of control individual ID SS_105. SNPs within the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position in ST30 reference (MRSA252; accession number: BX571856). NS– non-synonymous; S- synonymous; I- intergenic; *- denotes premature stop codon.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
124311	T->G	NS	SAR0114 *spa* immunoglobulin G binding protein A precursor	K->N	NC5
2202918	G->T	NS	SAR2140 *ilvD* dihydroxy-acid dehydratase	M->I	N1/2/3/4
1444841	A->G	NS	SAR1387 *femA* factor essential for expression of methicillin resistance	K->E	NC2/3
2311433	A->G	NS	SAR2242 *glmS* glucosamine--fructose-6-phosphate aminotransferase	I->T	NC2/3
17955	C->T	NS	SAR0013 membrane protein	A->V	NC2
372991	C->T	NS	SAR0326 membrane protein	G->S	NC2
1950542	C->A	NS	SAR1861 membrane protein	G->V	NC2
1663664	C->T	NS	SAR1590 membrane protein	A->T	NC1/4
2186507	T->G	STOP	SAR2126 *agrA* autoinducer sensor protein response regulator protein	L->*	NC5
605003	A->C	S	SAR0555 *kbl* putative 2-amino-3-ketobutyrate coenzyme A ligase		NC5
2510893	T->C	S	SAR2441 *tcaB* teicoplanin resistance associated membrane protein		NC5
2438295	C->T	S	SAR2368 ferrichrome-binding lipoprotein precursor		N1/2/3/4
706888	C->T	S	SAR0659 *fhuD* ferrichrome transport permease		NC3
252183	C->G	I	between SAR0216 lipoprotein and SAR0217 formate acetyltransferase		NC4
2566271	T->C	I	between SAR2490 hypothetical protein and SAR2491 acetyltransferase (GNAT) family protein		NC1/4

Table 4-7-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_105. Unique indels in the core genome differentiating sequenced colonies. Base position in relation to reference assembly used for mapping (105-NC3) and region/ gene are relative to position in ST30 reference genome (MRSA252; accession number: BX571856). (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
598890	NC1	(+) G	intergenic- putative promoter region of SAR2132 membrane protein	change in position of ribosomal binding site in relation to translational start site
45623	NC5	(-) a	intergenic between SAR1010 membrane protein and SAR1008 putative glycosyl transferases (pseudogene)	intergenic region; no obvious consequence; 29bp downstream of translational termination of SAR1010
483271	NC5	(+) T	SAR2591 *lysE* translocator	frameshift; SAR2591 lysE translocator
611579	NC5	(-) agt	SAR2119 *sdrH*	Within serine repeat region. Loss of single serine residue; no predicted consequence
2317291	NC1/2/3/4	84bp deletion	deletion within octapeptide repeat region of SAR0114 *spa* immunoglobulin G binding protein A precursor	truncated protein

4.8.3.3 Control study participant SS_149

This 6-year-old girl's colonising population represents the most diverse observed across all 18 controls that is also reflected in the estimated age of the carriage population of 24 months (Figure 4-5). The five colonies were differentiated by 17 SNPs, which is also the maximum genetic distance between any two colonies. Illustrated by the phylogeny (Figure 4-8) colony 4 is placed basally, separated from the remaining 4 colonies by a long branch. The lack of observed intermediate variation is might be suggestive of a well-established population. This raises the possibility of the population having been shaped by a selective sweep. However this may also reflect insufficient sampling to observe other intermediaries. Of the other 4 colonies a single SNP differentiates them, with NC3/2/and 5 being identical (Table 4-8-1). Two intergenic insertions were identified of which both are present in the basal NC4 (Table 4-8-2).

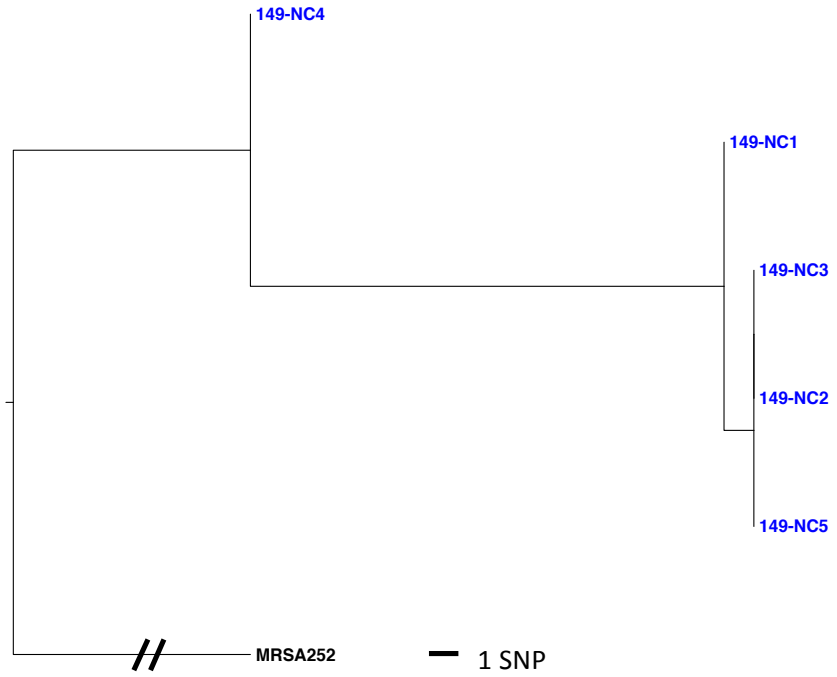

Figure 4-8 Maximum likelihood core SNPs tree for control study ID SS_149. Tree of 5 sequenced nasal colonies. Individual was colonised by an ST30 strain. Tree rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table 4-8-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_149. SNPs in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in ST30 reference (MRSA252; accession number: BX571856). NS– non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
1001282	A->C	NS	SAR0957 *appD* putative oligopeptide transport ATP-binding protein	I->L	NC4
1039472	C->T	NS	SAR0992 putative protease	A->V	NC4
1568899	G->A	NS	SAR1475 *aroA* 3-phosphoshikimate 1-carboxyvinyltransferase	P->S	NC4
2445203	G->A	NS	SAR2377 *ureG* urease accessory protein	G->S	NC4
590240	T->A	NS	SAR0546 hypothetical protein	I->N	NC1/5/2/3
671194	G->T	NS	SAR0623 membrane protein	D->Y	NC1/5/2/3
961489	C->T	NS	SAR0924 glucose-6-phosphate isomerase	R->C	NC1/5/2/3
1477613	T->C	S	SAR1420 hypothetical protein		NC4
2344042	A->G	S	SAR2262 UTP--glucose-1-phosphate uridylyltransferase		NC4
690035	A->G	S	SAR0643 ABC transporter ATP-binding protein		NC3/2/5
401664	C->T	S	SAR0355- Cys/Met metabolism PLP-dependent enzyme		NC1/5/2/3
2855974	T->C	S	SAR2753 lipase precursor		NC1/5/2/3
604696	C->A	I	between SAR0554 putative peptidase and SAR0555 2-amino-3-ketobutyrate coenzyme A ligase		NC4
1111893	C->T	I	between SAR1065 putative polypeptide deformylase 2 and SAR1066 lipoprotein		NC4
2810721	A->G	I	between SAR2714 *arcA* arginine deiminase and SAR2715 arginine repressor family protein		NC4
657126	T->A	I	between SAR0607 membrane protein and SAR0608 aldo/keto reductase family protein		NC1/5/2/3
2018634	A->C	I	between SAR1936 hypothetical protein and SAR1935 putative DNA repair exonuclease		NC1/5/2/3

Table 4-8-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_149. Indels differentiating sequenced colonies. Base position in relation to reference assembly used for mapping (149-NC5) and region/ gene are relative to position in ST30 reference genome (MRSA252; accession number: BX571856). (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
997044	NC4	(+) T	intergenic adjacent to SAR1263-hypothetical protein	intergenic region ; no predicted consequence
1424619	NC4	(+) T	intergenic between SAR1080 hypothetical protein and SAR1079	81bp upstream of SAR1079 in putative promoter region

	mntH divalent metal cation transporter

4.8.4 Potential relevance of diversity

The observed diversity and subsequent estimated time span of colonisation seemingly reflect the natural spectrum of *S. aureus* carriage observed in humans, which ranges from short-term or intermittent through to long-term. The levels of observed diversity in these individuals is in keeping with previous findings in long-term *S. aureus* carriers (Young et al. 2012), and therefore evidence of diversification in the individual as opposed to multiple acquisition events. The individuals exhibiting the greatest heterogeneity are therefore likely representative of extended periods of carriage, harbouring established *S. aureus* populations. Study of within host variation within these controls was used to systematically look for evidence of biologically relevant mutations, which may be indicative of the organism adapting to its respective host, as well as to the niche colonisation environment.

In the first instance, mutations that would be deleterious to function were searched for in each controls colonising population. Specific examples are given in section 4.8.3 and demonstrate loss of function mutations that will result in phenotypic heterogeneity within the individual. The clearest example of this is in control SS_105 where a nonsense mutation in the accessory global regulatory system (*agr*) component *agrA* is present in NC5 (Figure 4-7; Table 4-7-1). This premature stop mutation will result in expression of a truncated variant of the response regulator *agrA* and impacting on both the quorum sensing and overall virulence of these cells (Fowler et al. 2004; Le and Otto 2015). Within the context of the phylogeny this mutation has evidently arisen during a protracted period of carriage. Two further examples of truncating mutations were identified in this control's population, a frameshift mutation in the *lysE* family translocator (SAR2591) and a deletion within surface protein A (*spa*;SAR0114) (Figure 4-7; Table 4-7-1). The relevance of these mutations is less obvious. The *lysE* amino acid transporter

mutation, like the *agr* mutation, is found in NC5 is the single most diverse colony within the individual's population, also affected by the *agr* mutation. Whilst the surface protein A locates to a region of the *S. aureus* genome prone to variability given its repetitive structure, This gene has evidently varied in host with NC5 also having a SNP in the gene. However this deletion could potentially alter the wide-ranging functions of *spa*.

Secondly, mutations with the potential to alter or reduced function were assessed. Non-synonymous mutations in genes with functions in host survival and virulence were identified in these 3 controls. In control SS_039, two non-synonymous mutations were identified in NC3 in the gamma haemolysin precursor A (*hglA*) and *S. aureus* surface protein H (*sasH*), respectively (Figure 4-6; Table 4-6-1). Both are virulence factors associated with subversion of the host immune response by leucocyte cytolysis (*hglA*) or survival of the oxidative burst in neutrophils (*sasH*) (Thammavongsa et al. 2015) and with deletion mutants having reported impact on disease phenotype observed in murine models (Thammavongsa et al. 2009; Malachowa et al. 2011). These results suggest a trend in the accumulation of mutations that could potentially reduce overall virulence of this colonising strain, however whether this is functionally relevant is not known.

Finally, as another strategy to identify evidence of genetic adaptation to the host, mutations in promoter regions that would potentially alter gene expression, rather than ablating function, were looked for. In control SS_149, there was an insertion in an intergenic region upstream of the divalent metal cation transporter *mntH,* with potential subsequent disruption of the promoter for this gene (Table 4-8-2). Expression on *mntH* is believed to be an adaptive mechanism used by *S. aureus* to obtain manganese during active infection and bypassing the host's deliberate limitation of supply of this ion as a means of killing the organism (Kehl-Fie et al. 2013). Two previous population level studies have identified homoplastic mutations in this region of the genome of the ST239 and EMRSA15 lineages alike (Harris et al. 2010; Holden et

al. 2013). The functional consequence cannot be determined from this study, but it does highlight a potential fine scale adaptive change(s) arising in this colonising population.

4.9 Discussion

This study has allowed an initial characterisation of within host diversity of *S. aureus* carriage in healthy children. This demonstrated similar colonisation dynamics to which have been reported in previous adult studies of both hospital and community locations.

Carriage and strain prevalence in healthy children

Asymptomatic carriage studies have tended to focus on the adult population, whereas children are as frequently, if not more, prone to carriage and persistent carriage in particular (Kluytmans et al. 1997). The overall nasal carriage rate in this study was 29.1%, which is within reported rates of 19.9 to 36.2% in the 3 to 19 year old population in Europe (Blumental et al. 2013; Heijer et al. 2013). Higher carriage rates may have been identified if oropharyngeal carriage was assessed with throat swabbing, which can occur in the absence of nasal carriage and with greater frequency in infants (Esposito et al. 2014; Berents et al. 2015). Strain characterisation in childhood carriage studies is also less frequent than in adult equivalents. The 18 controls samples selected for sequencing demonstrated that the carriage was primarily of CC30 isolates (61%), with remainder being CC15 (17%) or CC45/ 22/5/1 (5% each). These findings are also in keeping with a previous study of Belgian pre-school children (Blumental et al. 2013), and outside of Europe in a study at a Brazilian nursery (Rodríguez et al. 2014). Overall the strain level characterisation in this small paediatric study is similar to those previously reported in European adult populations (Melles et al. 2004; Monecke et al. 2009).

Diversity of colonising populations in healthy children

Diversification of *S. aureus* during nasal carriage has been reported in healthcare associated MRSA transmission (Harris et al. 2013; Paterson et al. 2015; Tong et al. 2015), as well as in community based adults (Young et al. 2012; Golubchik et al. 2013). The samples sequenced for this work were studied to gain insight into the genetic diversity of healthy community based children given that this has not been addressed by previous studies. Across the carriage populations of these 18 healthy children similar micro variation in the form of core genome SNPs and indels was found compared with healthy adults (Golubchik et al. 2013). This ranged from no detectable variation through to a maximum of 17 core genome SNPs and 5 indels within any one individual's sequenced colonies. These levels of diversity would therefore be in keeping with a single acquisition event with subsequent diversification in that individual.

Having observed the preponderance to carriage of ST30 isolates a comparison was undertaken of the diversity across the individuals colonised by these strains to determine if the observed heterogeneity might be the result of cross transmission. This revealed core genome SNP distances between isolates from any two ST30 colonised controls with a minimum of 385 to a maximum of 513. This was also within the reported ranges of SNP divergence in adults in the community colonised by CC30 isolates (Golubchik et al. 2013). This added further support to the conclusion that the diversity observed in each control had arisen within that host and was not due a shared transmission between any two controls.

From the asymptomatic carriers assessed in this work we did not identify any examples of in-host variation in the carriage of accessory genome components conferring resistance properties. This was despite 10 of the individuals having been estimated to have colonising populations of 2.8 months or older (Figure 4-5). The resistance gene content analysis was undertaken to assess the prevalence of clinically relevant resistance markers in this healthy paediatric population. In host changes in carriage of AMR genes have been demonstrated in

hospital based populations (Lindsay et al. 2012), where widespread use of antimicrobials in the patient population would have influence on both the uptake and upkeep of such genes by the nosocomial *S. aureus* population. Therefore it is possible this type of variation wasn't observed in this study because we have selected healthy community based children, who are not being exposed to antimicrobials. Variability in the carriage population of other MGEs such as phage was not specifically assessed for, but which has previously been shown to occur during colonisation (Lindsay et al. 2012; Golubchik et al. 2013), but were not looked for in this study. Analysis of the indels however, did not reveal any examples that would have been indicative of phage loss or gain events.

Single strain colonisation

Multiple colony assessment per individual was incorporated into the study design to allow detection of co-colonisation, in addition to characterising diversity. Previous characterisation of isolates of community nasal carriers with *spa* typing showed that between 3 and 12 colonies were required to accurately identify co-colonisation, and showed an overall point prevalence of co-carriage of between 3.4 and 5.8% (Votintseva et al. 2014). No instances of co-colonisation were detected in the samples from these 18 children. For each control 5 colonies were sequenced, meaning there is potential for co-colonisation to have been present and but not detected. However from the perspective of time and cost feasibility extending to 12 colonies was not practical for this work. Of 4 examples of studies using WGS of 10 or more colonies from nasal carriers, including over protracted periods of carriage (Young et al. 2012; Golubchik et al. 2013; Paterson et al. 2015; Tong et al. 2015), there has been only a single instance of mixed strain carriage reported (Paterson et al. 2015). This might suggest that co-colonisation is potentially less frequent than previously reported albeit in the context of differing strain typing methods. When co-colonisation does occur it is hypothesised to represent a new transmission event (Didelot et al. 2016). Taken together the findings of single

strain colonisation and similar levels of in host genetic diversity indicate that the colonisation dynamics of this organism in healthy children are the same as previously shown in adults

Estimated age of the colonising* S. aureus *populations carried by healthy children

Using the diversity levels to estimate the duration of carriage in these children demonstrated a range from no diversity and hence a likely recent acquisition event, to a diversity suggesting a maximum age of the colonising population of approximately 24 months. These calculations are based upon an estimated base substitution rate (1.6 $x10^{-6}$ SNPs/site/year), which falls within ranges calculated for multiple *S. aureus* lineages (Uhlemann et al. 2014a). The rate chosen is similar to that derived from serial time point sampling in individuals. This study reported base substitutions rates in two patients over prolonged carriage periods of ST15 and 30 strains mutation rates of 1.87 $x10^{-6}$ SNPs/ site/ year (Young et al. 2012). The observed diversities and subsequent ages of the carriage populations inferred from them may reflect the carriage states of intermittent to long-term shown in experimental colonisation in humans (van Belkum et al. 2009b). In this work, persistent carriers were shown to maintain colonisation for 154 days (5.5 months) in comparison to intermittent carriers who were colonised for 14 days (van Belkum et al. 2009b).

Limited evidence suggestive of adaptation

Similar to the study by Golubchik *et al*. (2013) was limited evidence of adaptive mutations in this paediatric population during colonisation. There were examples of non-synonymous and truncation mutations in both surface proteins such as *spa* (Table 4-7-1), and toxins such as *hglA* (Table 4-6-1). These may represent antigenic loci and genetic alteration of could down regulate the potential to elicit a host response that would lead to clearance (Didelot et al. 2016). These are similar to the gene types identified in the carriage diversity work by (Golubchik et al. 2013). In addition, we identified a single instance of a premature stop mutation in *agrA* during carriage was identified (Figure 4-7; Table 4-7-1). The colonisation

population in this child was estimated to have arisen 15.7 months previously, which would be in keeping with longer-term carriage. In disease there would be a potential survival benefit conferred by gaining such a mutation because of reduced host clearance. This may also be the case in colonisation, where gaining a mutation such as this may allow switching to intracellular persistence (Tuchscherr et al. 2011) and hence prevent clearance.

The potential for functional repercussions of mutations accrued during carriage in the host are obvious when they arise in protein coding regions. However the effect of intergenic mutations, if any, is much less clear. An obvious exception to this will be if mutations affect regulatory regions such as promoters regions of for the adjacent genes. Clinically relevant examples of this type of mutation are in the multi-drug efflux pump *norA*, or the enoyl-acyl carrier protein reductase enzyme, *fabI* where promoter mutations result in quinolone and tricolsan resistance respectively (Kaatz et al. 2005; Grandgirard et al. 2015). In this control population there were two examples of mutations in regions upstream of putative promoters with the potential to affect the promoter region. In control SS_039 there was a three base deletion 65bp upstream of translational start site of a putative membrane protein (locus ID SAR2135; Table 4-6-2). Within samples from SS_149 there was a single base insertion 81bp upstream of the *mntH* divalent metal cation transporter (locus ID SAR1079; Table 4-8-2). It is unclear whether these had any functional consequences. They are of interest as no other examples were found in the controls, and they arose in individuals estimated as having colonising populations that arose 10 months or more before the sampling date, which raises the question of whether the mutations in SS_039 or SS_149 could be signalling subtle genetic modifications as a form of adaptation.

In SS_149 the insertion was more interesting because of its position in a region where homoplastic mutations have been identified in endemic *S. aureus* lineages by two previous studies (Harris et al. 2010; Holden et al. 2013). This suggests this region of the genome is under

selective pressure on a wider scale. The *mntH* gene is of relevance because it functions to allow *S. aureus* to sequester manganese to bypass vertebrate host restriction of the availability of the ion to kill pathogens (Hood and Skaar 2012; Kehl-Fie et al. 2013). Knockout mutants of *mntH* have not shown significant phenotypic changes in murine abscess models, but it is believed that in combination with another metal ion uptake system, MntABC, to be important for the organisms ability to withstand oxidative stress during infection (Horsburgh et al. 2002; Kehl-Fie et al. 2013). As discussed in the introduction chapter *P. aeruginosa* has been shown to utilise mutations in the promoter region of a haemoglobin scavenging system to its survival benefit in the host (Marvig et al. 2014).

4.10 Concluding remarks

This analysis has shown that healthy community based children are colonised with similar frequency to the adult population and by similar *S. aureus* strains. The genetic diversity observed in clonal *S. aureus* population colonising these 18 children is also similar to that previously described in adult carriage and transmission studies. These results provide a basis of comparison for *S. aureus* carriage in AE disease.

5 *Staphylococcus aureus* associated with Atopic Eczema prospective case study

5.1 Introduction

Decades of clinical studies have indicated a link between *Staphylococcus aureus* and the pathogenesis of atopic eczema (AE). Affected individuals are characteristically prone to colonisation by the organism, with reported rates of carriage exceeding 90% (Leyden et al. 1974; Hauser et al. 1985). Disease severity has been shown to correlate with the bacterial load and the number of body sites which are colonised (Williams et al. 1990; Lomholt et al. 2005; Tauber et al. 2016). Increasingly, there is a shift towards also understanding how microbial dysbiosis as a whole contributes to the development of AE and subsequently disease flares. Metagenomic studies of skin microbiota in AE have shown that changes in population level microbial communities are significantly associated with disease activity in AE (Kong et al. 2012). Importantly this study reaffirmed that *S. aureus* specifically is intrinsically linked to increasing severity, however the main limitation of the metagenomic approach is the lack of discriminatory power to resolve this *S. aureus* population beyond the species level.

The DNA library preparation and whole genome sequencing for this study was kindly carried out by Dr Kerry Pettigrew, Infection Group, University of St Andrews.

5.2 Aims and objectives

This study aimed to investigate the micro-epidemiology of colonisation by S*. aureus* in children with active AE, by sampling across multiple body sites including carriage and disease affected skin in each case. WGS of multiple *S. aureus* colonies per body site in each individual were then used as a high resolution-typing tool to characterise the colonising populations of children with moderate to severe AE. The initial comparison was at strain level between body sites looking for evidence of strain heterogeneity in the colonising backgrounds. Extending from this was a

detailed assessment to look for evidence of genetic diversity arising during colonisation. Each participant's samples were then assessed to look for evidence of adaptive changes arising during carriage within the host, which may have biological relevance to the disease-promoting potential of this organism in the context of AE.

5.2.1 Recruitment

Given that this study was exploratory in nature an estimate of sample size was not undertaken. Instead a target sample size similar to previous in-depth microbial sampling studies in AE was used. This was with the intention of recruiting 10 children to the study.

Eleven children with AE were sequentially recruited to this study through the Paediatric eczema clinic at Ninewells Hospital between February and October 2015. Case study participants were selected specifically on the basis of having symptoms and examination findings in keeping with active eczema on the day of attendance at clinic. Additionally they had not received any antimicrobial or antiseptic therapy in the 4 weeks prior to the clinic, which might impact upon their *S. aureus* carriage. Significant delays incurred during recruitment because of widespread prescription of topical antiseptics to children with AE prior to their referral to dermatology. It importantly highlighted the extensive use of these therapies outwith current recommended guidance on their use (National Institute for Health and Care Excellence 2007). As a result of these issues, the anticipated recruitment period has to be extended from 8 weeks to 6 months, therefore limiting the opportunity to extend the subsequent study.

5.2.2 Sampling method and rationale

To investigate the colonisation of children with AE disease flares, skin swabs were obtained from 5 body sites. These included: the nostril (as the presumed primary carriage site), two

clinically unaffected skin sites, and two actively inflamed skin sites. Each eczema site was sub-sampled, by swabbing the lateral and medial edges of the area of inflamed skin. This sampling methodology allowed comparison between the presumptive reservoir and disease sites to examine whether there was cross transmission between the sites, or if colonisation at disease sites varied from nasal carriage.

Swabs obtained from each of the 5 body sites sampled were then processed as described (Material and Methods Section 2.3.2). From each positive swab, representative of a single body site, 5 colonies were randomly selected for DNA extraction and WGS. Multiple colonies were selected for comparison to determine if there was evidence of diversity in the colonising population in AE, and from multiple body sites to determine if colonisation was by the same strain across these sites. Colonies from sub-sampling within a single eczema site were included to specifically assess if there was genetic heterogeneity arising within the colonising population at inflamed skin sites. Five colonies were chosen at the time of commencement of this study to provide a cost-effective survey of diversity.

5.3 Clinical phenotype of recruits

Of 11 children with AE recruited to this study, 2 were not colonised by *S. aureus* in any of the five body sites sampled and were subsequently excluded from the analysis (Study IDs PSAE007 and PSAE011). The clinical characteristics of the remaining 9 recruits are shown in Table 5-1. Cases were selected on the basis of having moderate to severe AE. As a recommended standard of reporting AE severity their disease was scored by using the EASI classification (Schmitt et al. 2014). This EASI scoring method takes into account body surface area affected by AE and the severity of inflammation at disease sites amongst other features. Parent reported severity of AE was also taken into consideration, including symptoms of disturbed sleep, observed scratching and lack of response to treatment. Patient 5 had locally severe

disease only, but severe symptoms and treatment failure, hence there is not a reported EASI score.

Table 5-1 Atopic eczema case characteristics. Case numbers are equivalent to last digits of study ID for each patient. (y- years, m- months, N/A- not applicable)

Case	Age	Sex	EASI Score	Atopic disease (other than AE)	Other inflammatory skin disease	Antimicrobial therapy (< 4 weeks previously)
1	4 y	M	19.2	Hayfever	Nil	Nil
2	3 m	F	12.8	Nil	Nil	Nil
3	2 y 8 m	M	23.6	Food allergy	Nil	Nil
4	5.5 m	F	25	Nil	Nil	Nil
5	2 y 3 m	M	N/A	Food allergy	Nil	Nil
6	4.5 m	M	23	Nil	Nil	Nil
8	2 y 6 m	F	18.4	Hayfever	Nil	Nil
9	6 m	F	21.3	Food allergy	Nil	Nil
10	8 m	F	37	Nil	Nil	Nil

5.4 AE disease severity and S. aureus burden from clinical samples

Disease severity in AE has been shown to correlate to both the number of body sites colonised as well as the bacterial burden from lesional sites (Williams et al. 1990; Lomholt et al. 2005). Shown in Table 5-2 displays the disease grading based upon EASI score and the number of body sites found to be colonised by *S. aureus* in each patient.

Table 5-2 Correlation of disease severity and number of body sites colonised by *S. aureus* in AE cases. N- indicates that nasal colonisation was detected.

Study ID	EASI score	Disease severity	Number of colonised body sites
PSAE001	19.2	Moderate	5^N
PSAE002	12.8	Moderate	1
PSAE003	23.6	Severe	2
PSAE004	25	Severe	5^N
PSAE005	N/A	Locally severe	1
PSAE006	23	Severe	2
PSAE008	18.4	Moderate	4^N
PSAE009	21.3	Severe	2
PSAE010	37	Severe	5^N

Three cases were found to be colonised at all the sampled sites, all of which had moderate or severe disease. A single case was colonised at 4 body sites, whilst the remainder were colonised at two or fewer sites. Patients 1, 4, 8 and 10 had the greatest bacterial burdens, and were all affected by moderate to severe disease (Table 5-3). Patients 3 and 6 despite severe disease were colonised at only 2 sites and with less bacterial burden than other severe cases, which may be a reflection of their disease phenotype of widespread but superficial eczema. This refers to the degree of inflammation occurring within the skin, and the depth of the cutaneous layers affected. Superficial eczema in general is considered to be more acute, and with less dermal inflammation. With more chronic disease there is accrual of inflammation transcending to the dermal level, which is evident clinically in the form of lichenification (or exaggeration of skin markings and creases) (Leung 1995; Oranje et al. 2007). This feature was more prominent in the patients 1, 4, 8 and 10 and is a clinical sign of persistent scratching. This would be in keeping with previous reports that local, or lesional disease severity and hence lichenification, is associated with greater colony burden (Tauber et al. 2016).

Table 5-3 Bacterial burden in AE cases. CFU counts from 9 cases sampled across 5 body sites. Cells highlighted in grey indicate where colony count was recovered only by 18-hour enrichment. EASI (Eczema Area Severity Index) indicating disease severity. Sites sampled: N-nose, U1- unaffected site 1, U2- unaffected site 2, E1/2- lateral/ medial border of single eczema site, E3/4- lateral/ medial border of single eczema site.

Case	N	U1	U2	E1	E2	E3	E4
1	2.0×10^4	1.8×10^5	2.75×10^6	1.70×10^5	3.50×10^5	2.60×10^6	2.36×10^6
2	Negative	Negative	Negative	1.40×10^5	Negative	Negative	Negative
3	Negative	Negative	1.0×10^4	6.0×10^4	7.20×10^5	Negative	Negative
4	4.50×10^5	1.70×10^5	6.0×10^4	3.36×10^6	4.96×10^6	5.0×10^6	3.16×10^6
5	Negative	Negative	Negative	7.0×10^4	5.0×10^4	Negative	Negative
6	Negative	Negative	2.0×10^4	Negative	Negative	4.0×10^4	4.0×10^4
8	3.10×10^5	Negative	2.0×10^4	3.50×10^6	1.30×10^6	9.0×10^5	1.4×10^6
9	negative	Negative	Negative	1.0×10^6	1.30×10^6	Negative	3.0×10^4
10	1.44×10^6	8.0×10^4	Negative	1.0×10^5	3.0×10^4	6.4×10^6	7.4×10^5

5.5 Colonising strains

Up to 5 colonies were selected per colonised body site for sequencing. The numbers of colonies sequenced per case varied in accordance with the number of colonised sites where

available from the microbiological plate as well as the recovered bacterial burden. In each case a minimum of 10 and up to a maximum of 28 colonies were sequenced. There was variation in number of colonies sequenced per case because of variable number of sites identified as being *S. aureus* colonised in each individual.

Presented in Table 5-4 are the MLST profiles extracted from the assemblies of all sequenced colonies per case. This showed that colonisation in all cases was by a clonal population represented by a single ST with the exception of patient 8 (Study ID PSAE008). In this patient two distinct STs were found to co-colonise all body sites, ST123 and 45. In total, four of the nine (44.4%) cases were colonised by strains belonging to CC1 (ST1 and ST188). Otherwise isolates were derived across a further 6 clonal backgrounds, including single instances of ST30, ST45, ST5 and ST59 strains.

Table 5-4 Colonising strains in AE cases. Sequence type (ST), clonal complex (CC), and MLST profile is representative of all sequenced colonies for the individual. * Indicates case was found to be co-colonised by two distinct STs. (*arc*- carbamate kinase 1; *aroE*- shikimate dehydrogenase; *glpF*- glycerol kinase; *gmk*- guanylate kinase; *pta*- phosphate acetyltransferase; *tpi*- triosephosphate isomerase; *yqil*- acetyl conenzyme A acetyltransferase

Study ID	ST	CC	*arcC*	*aroE*	*glpF*	*gmk*	*pta*	*tpi*	*yqil*
PSAE001	188	1	3	1	1	8	1	1	1
PSAE002	59	59	19	23	15	2	19	20	15
PSAE003	1	1	1	1	1	1	1	1	1
PSAE004	1	1	1	1	1	1	1	1	1
PSAE005	30	30	2	2	2	2	6	3	2
PSAE006	5	5	1	4	1	4	12	1	10
PSAE008*	123	121	6	5	6	2	7	17	19
PSAE008*	45	45	10	14	8	6	10	3	2
PSAE009	2867	12	1	152	1	8	1	5	11
PSAE010	1	1	1	1	296	1	1	1	1

5.6 Resistome profile of AE associated strains

Children with AE are prone to disease exacerbations that are often attributed to *S. aureus* colonisation. Subsequently as a patient group they receive repeated courses of antimicrobial therapy. To further characterise the colonising populations of these AE cases the resistance

profile for each of the sequenced colonies was then assessed. This was also intended to look for evidence of in host diversification in the accessory genome to be detected.

Table 5- 5 demonstrates the resistance determinants common to all sequenced colonies for each case. This revealed two examples of in host variation of the accessory genome. In patients 1 and 5 (PSAE001/5) respectively there is evidence of variable plasmid carriage. In patient 1 (PSAE001), 21 of the 25 sequenced colonies were found to have a plasmid carrying the β-lactamase gene, *blaZ* (plasmid backbone similar to pSaa6159; accession number CP002115) such that the colonising population would be variably sensitive to penicillin. For patient 5, similar variance was noted between the 10 sequenced colonies, half of which were carrying a 27kb plasmid (backbone similar to SAP0194; accession number GQ900385.1) with the heavy metal resistance determinants *arsB/C, cadA* and *mco*.

β-lactamase carriage was found in 5 of the 10 colonising strains, from 9 cases. Determinants of fusidic acid resistance, a common therapeutic intervention in children with AE, were found with similar frequency. Three cases were carrying *fusC* (O'Neill et al. 2007), whilst a single case carried *fusB* (O'Neill et al. 2004). The core gene point mutation in codon 461 of *fusA* (translation elongation factor) is reported to confer high level resistance to fusidic acid (Chen et al. 2010) was found to be common to all 25 sequenced colonies from patient 1. Two cases were found to be colonised by strains carrying the *qacC* gene (Littlejohn et al. 1991), which confers diminished susceptibility to antiseptics such as benzalkonium chloride that are commonly prescribed for AE. The erythromycin resistance gene *ermC* (Catchpole et al. 1988) was found in a single case (patient 2; study ID PSAE002). No other core genome mutations conferring resistance were identified within this group.

Table 5-5 Resistance profiles of AE case isolates. Results per individual are representative of all sequenced colonies. Boxes indicated with – denote absence of resistance gene. Grey shading indicate resistome profiles of the cases where variability in resistance determinant carriage was identified within the clonal population. (Gene abbreviations are as defined in Materials and Methods section 2.6.9, Table 2-15).

Study ID	ST	*blaZ*	*fusB*	*fusC*	*ermC*	*qacC*	*cadD*	*arsB*	*arsC*	*arsC*	*cadA*	*mco*	*fusA* L461K
PSAE001	188	*blaZ*	-	-	-	-	*cadD*	-	-	-	-	-	*fusA*
PSAE001	188	-	-	-	-	-	-	-	-	-	-	-	*fusA*
PSAE002	59	-	-	-	*ermC*	-	-	-	-	-	-	-	-
PSAE003	1	-	-	*fusC*	-	-	-	-	-	-	-	-	-
PSAE004	1	*blaZ*	-	*fusC*	-	*qacC*	*cadD*	-	-	-	-	-	-
PSAE005	30	*blaZ*	-	-	-	-	-	*arsB*	*arsC*	*arsC*	*cadA*	*mco*	-
PSAE005	30	*blaZ*	-	-	-	-	-	-	-	-	-	-	-
PSAE006	5	*blaZ*	-	-	-	-	-	-	-	-	-	-	-
PSAE008	123	*blaZ*	*fusB*	-	-	-	-	-	-	-	-	-	-
PSAE008	45	-	-	-	-	*qacC*	-	-	-	-	-	-	-
PSAE009	2867	-	-	-	-	-	-	-	-	-	-	-	-
PSAE010	1	-	-	*fusC*	-	-	-	-	-	-	-	-	-

5.7 Genetic heterogeneity within individuals

5.7.1 Diversity of colonising *S. aureus*

The colonising populations from AE cases were all found to be clonal on the basis of MLST analysis. This led to the consideration that during AE disease exacerbations there is a clonal expansion of the *S. aureus* on the skin potentially stemming from a pre-existing population within the patient. To further explore these populations the genomic diversity of the isolates was characterised by assessment of individual SNPs, insertions and deletions and reconstruction of the phylogeny of each individuals colonising population.

In all cases the sequenced colonies were mapped to a reference genome of the same clonal complex, or where unavailable a draft assembly from one of that patient's own colonies. Accessory regions were masked and core genome SNPs were identified. The relative diversity in each host was then quantified by calculating the number of core genome SNPs per

sequenced colony. Each cases' samples were then also mapped to a self-assembly to identify unique indels which had arisen during carriage.

From this analysis it became apparent that there was a range of diversity across each of the AE cases, just as had been found in the nasal controls. Patient 2 (study ID PSAE002) had the lowest observed level of diversity at 0.07 SNPs per colony. This patient notably had the lowest disease severity (Table 5-2) and colonisation at a single body site. This low level of diversity reflects the single core genome SNP, which differentiated the 14 sequenced colonies. Increasing diversity was noted in individuals with moderate to severe disease (Figure 5-1). The overall greatest diversity was found in patient 5, affected by localised disease only, at 2.6 SNPs per colony.

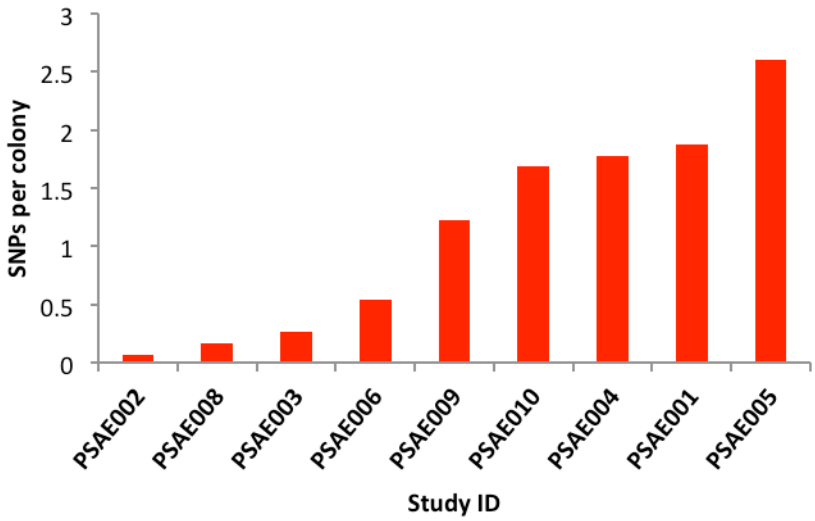

Figure 5-1 In-host diversity levels across AE cases. Diversity calculated as number of core genome SNPs/ sequenced colony/case.

5.7.2 Estimating the age of the colonising population of *S. aureus* in AE cases

The observed levels of diversity across the cases suggested that in some instances colonisation persists over prolonged periods of time. Utilising the maximum pairwise divergence between

colonies in each individuals population, divided by two, allowed an estimation of the age of the colonising population to be derived, as described in Chapter 4.

Table 5-6 Size of core genome from representative reference genomes and self assemblies used for mapping and SNP assessment. Estimated SNP rates for each lineage are based upon size of core genome. *No regions were excluded as accessory from self-assemblies and calculations were based on the total assembly size.

Reference genome/ assembly	Assembly size (bp)	Accessory genome regions excluded (bp)	Core genome size (bp)	Estimated SNPs/ genome/ month
MSSA476	2799802	172174	2627628	0.35
MRSA252	2902619	266437	2636182	0.35
CUHK_HK188	2807977	185559	2622418	0.35
CA347	2850503	223275	2627228	0.35
N315	2814816	205466	2609350	0.35
M013	2788636	129857	2658779	0.35
PSAE009_E4C2	2745212	N/A	2745212	0.36

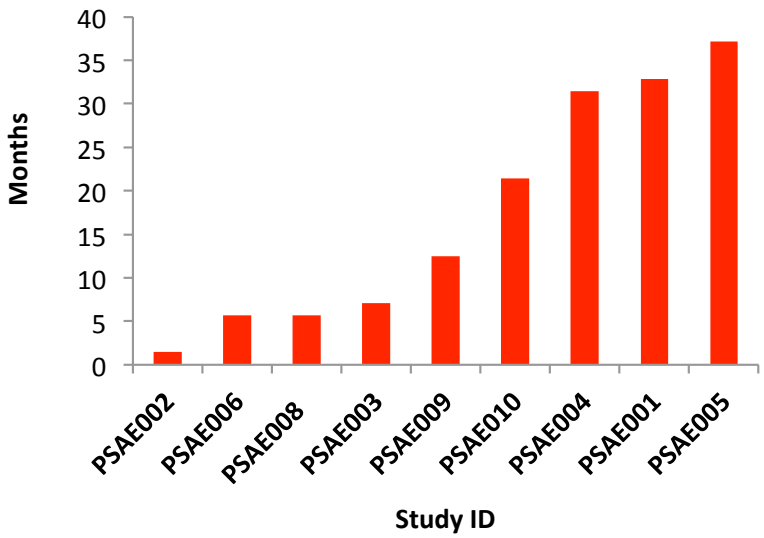

Figure 5-2 Estimated age of the colonising population in Atopic eczema cases. Time to most common recent ancestor based on maximum pairwise SNP distance observed in the sampled population, presented in months and is calculated on estimated mutation rate of 1.6 x10^{-6} base substitutions/site/year, on basis of core genome size presented in Table 5-6.

This analysis suggested colonising populations in cases had arisen between 1.4 to 37 months previous to the sample date (Figure 5-2). In patient 2, 3, 6 and 8 the age of carriage population was estimated at between approximately 1.4 and 7 months. The remaining patients had

populations with estimated ancestral origins of between 12.5 months previously for patient 9, and to a maximum of 37 months for patient 5. In the case of patient 5, there was parental reporting of eczema symptoms that has been on-going for more than 5 months at the time of the clinic appointment, with failure to improve with emollients or extended topical steroid treatment. Similar histories were obtained from parents of patients 1 and 4, where their eczema had been active for months prior to their assessment at clinic.

5.7.3 Diversity within individual AE cases

The diversity in each AE case was then characterised and examined in the context of their phylogeny. The identified core genome SNPs were used to reconstruct a phylogeny. Core genome SNPs were reconstructed back on to the phylogeny to allow each mutation differentiating colonies or parts of the colony population within an individual to be defined.

Across the nine AE cases varying degrees of in host diversity were observed as previously illustrated by the pairwise SNP distance analysis. Colonies within any individual's sampled population were found to be separated by a minimum of 1 core genome SNP in the instance of patient 2, through to a maximum of 26 core genome SNPs in patient 5. Similar levels of diversity have been described in nasal carriage populations (Golubchik et al. 2013; Tong et al. 2015), but not previously assessed in cutaneous colonisation in AE.

For each case studied, there were examples of in-host variation with potentially relevant clinical implications. The following results give specific examples of these findings. The phylogeny, core genome SNPs and indels identified in the remaining cases are presented within the supplementary data appendix (Appendix B).

5.7.3.1 Clonal expansion in colonising population and self-transmission

Representative of one of the most diverse populations in the AE cases was patient 1. A total of 25 colonies were sequenced from this 4-year-old boy with moderate eczema. Within these colonies derived across 3 body sites, were 47 core genome SNPs, with a maximum SNP distance between the two most diverse colonies being 23. A total of 5 unique indels were also identified as having arisen during the colonisation period in this child. The SNPs and indels differentiating the colonies within this child's sequenced samples are presented in Tables 5-7 and 5-8.

The phylogeny of this child's colonies demonstrates the diversity of the colonising population derived from sampling of 4 body sites. Three colonies can be seen to position basally in the phylogeny, separated from the remaining 22 colonies by a long branch (without evidence of recombination), representing a genetic distance of 22 core genome SNPs (Figure 5-3). Within the 22 colonies of the clade, there are 3 pairs of colonies identical at core genome levels, but separated from the other colonies by a single SNP. Five colonies from an unaffected skin site are identical at core genome levels, but sitting within a clade with a colony derived from an eczema site.

This case was one of only 4 identified as nasally colonised in this study. The first observation is that colonies derived from all of the sampled sites can be seen interspersed in the phylogeny. Nasal colonies are seen basally within this tree alongside a colony derived from an eczema site (E2) (Figure 5-3). The nasal colonisation would be the presumptive reservoir, but it cannot be proven that this was the initial source. Nasal colonies can otherwise been seen intermingling with eczema site colonies which demonstrates sharing of the colonising populations between the nose and other body sites. The lack of observed genetic diversity in the unaffected skin site

colonies suggests that they are a more a more transitory population at this site. They are likely representative of a recent colonisation event resulting from direct spread across the body.

The resistome profiling of this young boy's colonising population also demonstrates diversity in its antimicrobial sensitivity. Colonies in the phylogeny NC2/ E2-ic2/ N-18-c3 and N-18-c1 were found to be lacking in carriage of a β-lactamase in comparison with the other colonies. Their basal position is suggestive of their ancestral status and perhaps indicates that the plasmid (backbone similar to pSaa6159; accession number CP002115) conferring *blaZ* and *cadD* (Figure 5-5) was gained during the period of carriage. The majority of the colonies with this plasmid are extra-nasal, which raises the possibility that acquisition of the plasmid provided an advantage to the colonising strain at these sites.

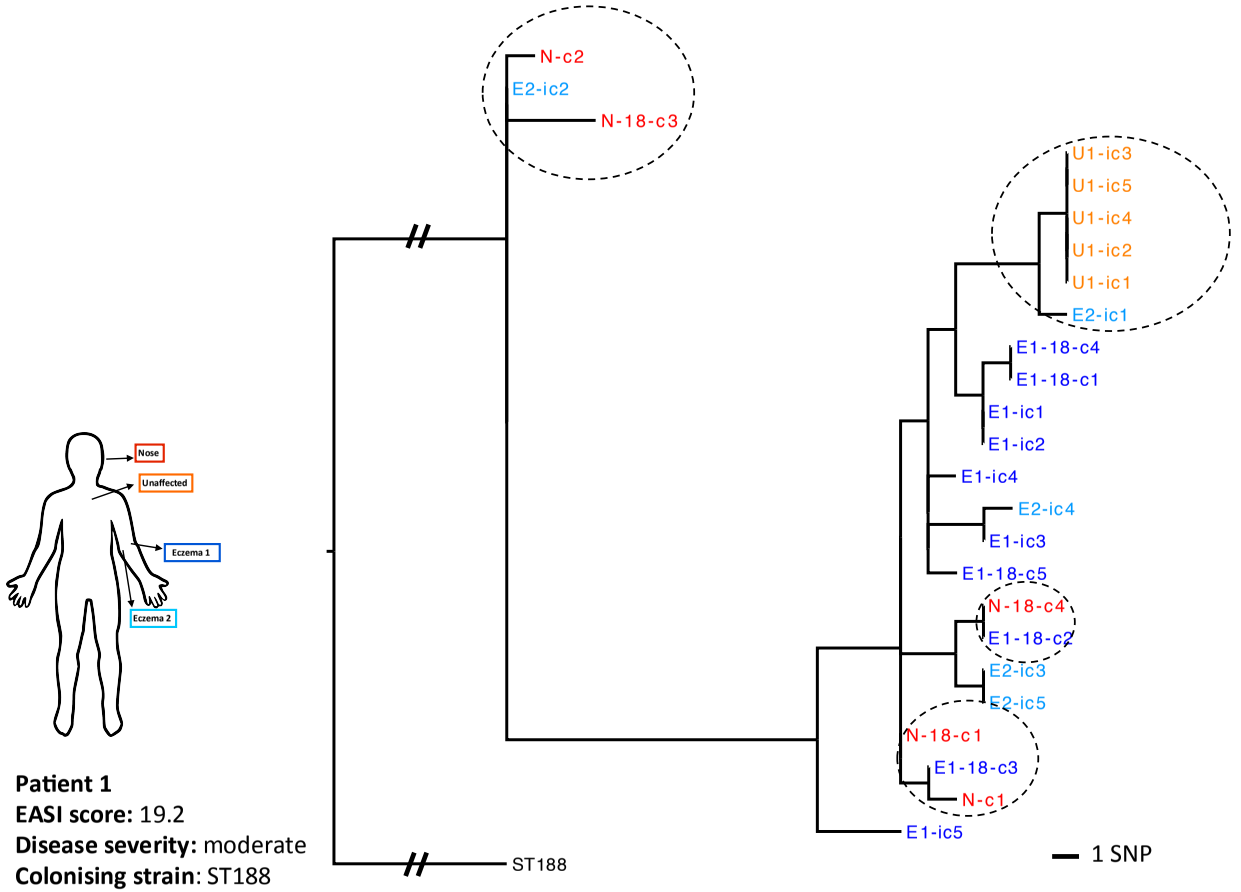

Figure 5-3 Maximum likelihood core SNPs for Patient 1 (study ID PSAE001). Tree demonstrates the phylogenetic relationships between the 25 colonies from 3 body sites. Colonies were obtained by both pre- (labeled ic- initial colony) and post-enrichment (labeled -18) culture methods. Body diagram shows sampling site, branch label coloring corresponds to body site colony was obtained from. Perforated circles indicate evidence of transmission between body sites. Branch labels: E- eczema site, U- unaffected site, N- nose, C- colony number from body site. Tree is rooted using ST188 reference. Scale bar is approximately 1 SNP (not applicable to root branches with strikethrough). (EASI- Eczema Area Severity Index).

Table 5-7 Within host-heterogeneity identified in Patient 1 (study ID: PSAE001). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in ST188 reference (CUHK_HK188; accession number: JFFV00000000) with annotation transferred from MSSA476. NS- non-synonymous; S- synonymous; I- intergenic.

Patient 1 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
123172	C->T	NS	SAS0138 *capO* capsular polysaccharide synthesis enzyme	H->Y	NC2
1397002	C->T	NS	SAS1359 putative membrane protein	R->H	NC2
278320	A->G	NS	SAS0279 conserved hypothetical protein	K->Z	N18-c3, E2-ic2, E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
862339	A->G	NS	SAS0792 conserved hypothetical protein	A->T	N18-c3, E2-ic2, NC2
1590300	A->T	NS	SAS1522 *lepA* putative GTP-binding protein	L->M	N18-c3
63535	G->C	NS	SAS0085 *spa* immunoglobulin G binding protein A precursor	A->G	E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
121239	C->T	NS	SAS0137 *capN* capsular polysaccharide synthesis enzyme	T->I	E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
1952977	C->T	NS	SAS1861 MHC class II analogue (pseudogene)	V->I	E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5

Patient 1 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
2477451	T->A	NS	SAS2439 TetR family regulatory protein	D->E	E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
352155	C->T	NS	SAS0354a hypothetical protein	A->V	E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4
1151746	A->T	NS	SAS1136 *pyrA* putative carbamoyl-phosphate synthase	H->L	E2-ic5, E2-ic3, E1-18-c2, N18HR-C4
2337850	G->A	NS	SAS2311 *hlgC* gamma-hemolysin component C precursor	A->T	E2-ic5, E2-ic3, E1-18-c2, N18HR-C4
2569251	T->A	NS	SAS2523 zinc metalloproteinase aureolysin precursor	T->S	E2-ic5, E2-ic3
1406572	G->A	NS	SAS1372 cvfC conserved virulence factor C	A->V	N-c1
1358058	C->G	NS	SAS1322 *oppC2* putative oligopeptide transport system permease	A->P	E1-ic3, E2-ic4
671963	T->G	NS	SAS0650 ABC transporter ATP-binding protein	F->V	E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3
262422	C->T	NS	SAS0257 putative exported protein- homology to staphyloxanthin biosynthesis protein	A->T	E1-18-c1, E1-18-c4, E1-ic2, E1-ic1
203697	C->T	NS	SAS0207 *fadA* (putative thiolase)	R->K	E1-18-c1, E1-18-c4
411773	A->T	NS	SAS0411 hypothetical protein	L->F	E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3
934787	A->G	NS	SAS0860 transport system extracellular binding lipoprotein- peptide transport activity	K->E	E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3

Patient 1 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
132130	C->T	NS	SAS0149 putative transport system permease NitT/ TauT family transport permease protein	P->L	E1-18-c5
2546025	G->A	NS	SAS2505 precorrin-2 dehydrogenase	A->V	E1-ic4
665535	G->A	NS	SAS0644 putative acetyletrasnferase	P->L	E1-ic5
2059855	G->T	NS	SAS2032 putative acetyltrasnferase	Q->K	E1-ic5
2417876	G->A	NS	SAS2387 Fibronectin binding protein A	P->L	E1-ic5
1716227	A->G	S	SAS1639 conserved hypothetical protein		NC2
547760	A->G	S	SAS0521 *sdrE* bone sialoprotein-binding protein		N18-c3, E2-ic2, E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
547802	C->T	S	SAS0521 *sdrE* bone sialoprotein-binding protein		N18-c3, E2-ic2, E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
547805	T->C	S	SAS0521 *sdrE* bone sialoprotein-binding protein		N18-c3, E2-ic2, E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
1716227	A->G	S	SAS1639 conserved hypothetical protein DNA methyltransferase		N18-c3, E2-ic2, NC2
1298571	G->A	S	SAS1268 putative homoserine dehydrogenase		N18-c3
1413994	G->A	S	SAS1377 *ebH*		E1-18-c2, N18HR-C4

Patient 1 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
2485414	A->T	S	SAS2446 putative aminotransferase		E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4
1310226	T->C	S	SAS1279 *lexA* DNA damage-inducible repressor		E2-ic1
487186	G->A	I	Not annotated; corresponds to rRNA region in MRSA252		N18-c3, E2-ic2, E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
1925292	A->C	I	Between SAS1831 (adenylosuccinate lyase) and SAS1832 (staphopain protease)		N18-c3, E2-ic2, E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
481415	G->T	I	Intergenic between SAS0474 (lysyl-tRNA synthetase) and SAS0475 (GntR family regulatory protein)		E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4
680435	T->C	I	Between SAS0659 (conserved hypothetical protein) and SAS0660 (fluoroquinolone resistance protein)		E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4
1165397	C->A	I	Between SAS1146 (primosomal protein n') AND SAS1147 (putative lipoprotein)		E1-ic3, E2-ic4
1165397	C->A	I	Between SAS1146 (primosomal protein n') AND SAS1147 (putative lipoprotein)		E1-ic3, E2-ic4

Patient 1 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
791348	G->T	I	Between SAS0754 (putative exported protein) and SAS0755 (putative exported protein)		U1-ic3, U1-ic5, U1-ic4, U1-ic2, U1-ic1
395694	A->G	I	Between SAS0395 (putative restriction and modification system specificity protein) and SAS0396 (exotoxin)		NC2
1765608	C->T	I	Between SAS1677 (putative peptidase) and SAS1678 (putative exported protein)		NC2
1765608	C->T	I	Between SAS1677 (putative peptidase) and SAS1678 (putative exported protein)		N18-c3, E2-ic2
1108850	T->A	I	Between SAS1094 (putative membrane protein) and SAS1095 (putative membrane protein)		N18-c3
1765601	C->G	I	Between SAS1677 (putative peptidase) and SAS1678 (putative exported protein)		E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
2197051	C->T	I	Between SAS2173 (putative inosine-uridine preferring nucleoside hydrolase) and SAS2174 (putative ferrichrome-binding lipoprotein precursor)		E2-ic5, E2-ic3, E1-18-c2, N18HR-C4, E1-18-c3, N-c1, N18-c1, E1-ic3, E2-ic4, E1-18-c1, E1-18-c4, E1-ic2, E1-ic1, E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3, E1-18-c5, E1-ic4, E1-ic5
1309460	A->T	I	Between SAS1278 (hypothetical protein) and SAS1279 (DNA damage-inducible repressor)		E1-18-c3, N-c1
544345	C->A		Between SAS0520 (LPXTG surface protein) and SAS0521 (bone-sialo protein)		E2-ic1, U1-ic1, U1-ic2, U1-ic4, U1-ic5, U1-ic3

Table 5-8 Within host-heterogeneity identified in Patient 1 (study ID: PSAE001).
Unique indels in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to reference assembly used for mapping PSAE001_E2_IC1. Region and gene are in position to ST188 reference (CUHK_HK188; accession number: JFFV00000000) with annotation transferred from MSSA476. (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
2571319	E1 18 C1, E1 18 C4	(+) A	Intergenic region between SAS0380 (conserved hypothetical protein) and SAS0381 (conserved hypothetical protein)	No obvious consequence
908129	NC2, E2IC2, N18 C3	(+) GT	Intergenic region between SAS1107 (putative DNA-binding protein) and SAS1107a (antibacterial protein)	No obvious consequence
365382	NC2, E2IC2, N18 C3	(-) ta	In ST188 gff deletion present in ST188_02373- Sau3AI Type 2 restriction modification system	Possible functional change
1693390	E1IC5	(+) A	Integrenic region between ST188_00677 (Predicted permeases,carboxylate/amino acid/amine transporter) and ST188_00678 (Nrdi,hypothetical protein,ribonucleotide reductase stimulatory protein,nrdI protein,NrdI Flavodoxin like)	No obvious consequence
1370737	N18 C3	(-) t	Intergenic region between ST188_00003 (Transcriptional regulator,HTH-type transcriptional regulator immR,hypothetical protein,) and ST188_00004 (protein of unknown function DUF3147)	No obvious consequence

5.7.3.2 Genetic adaptation of S. aureus during colonisation

A total of 28 colonies were sequenced from patient 4, a 5 and half-month old female with severe AE. Amongst the colonies derived across 5 body sites were a total of 50 unique core genome SNPs and 6 indels. The two single most distantly related colonies in this clonal population were distinguished by 22 core SNPs.

From the phylogeny (Figure 5-4) two colony sub-populations can be seen within this patient. The most basal isolates represented in the top half of the tree, are distinguished from a

separate clade by a branch representing a distance of 5 core genome SNPs. Within the basal isolates, 3 colonies are genetically identical, all from unaffected site 2 (taxon labels coloured in orange in Figure 5-4). The rest of these basal isolates are genetically distinct by distances varying from between 1 core SNP to 5 in the case of colony E2-C5. Within the clade separated by 5 SNPs from the rest of the colonies, 5 are identical at core genome level. The colonies are otherwise separated by between 1 and 4 core SNPs.

The assessment of the variation present within this child's colonies revealed two separate function changing mutations in the accessory global regulatory (*agr*) system component *agrA*. The homoplastic mutations of this gene are highlighted in the phylogeny, present within both parts of the two colony sub-populations. A single colony in the basal population, E2C5, was found to have a nonsense mutation in *agrA* (Table 5-9). Whilst all 18 colonies within the distinct clade had a single base deletion in *agrA*, giving rise to a truncation (Table 5-10). Both of these mutations would be expected to ablate the functioning of the *agrA* response regulator, as a key component of this virulence regulatory system. As such the overall virulence and quorum sensing capacity of the colonies would be impacted upon (Fowler et al. 2004). The presence of two distinct mutations in this gene, across the population is an example of convergent evolution, which is strongly suggestive of this mutation having provided a selective advantage to surviving colonisation in this individual.

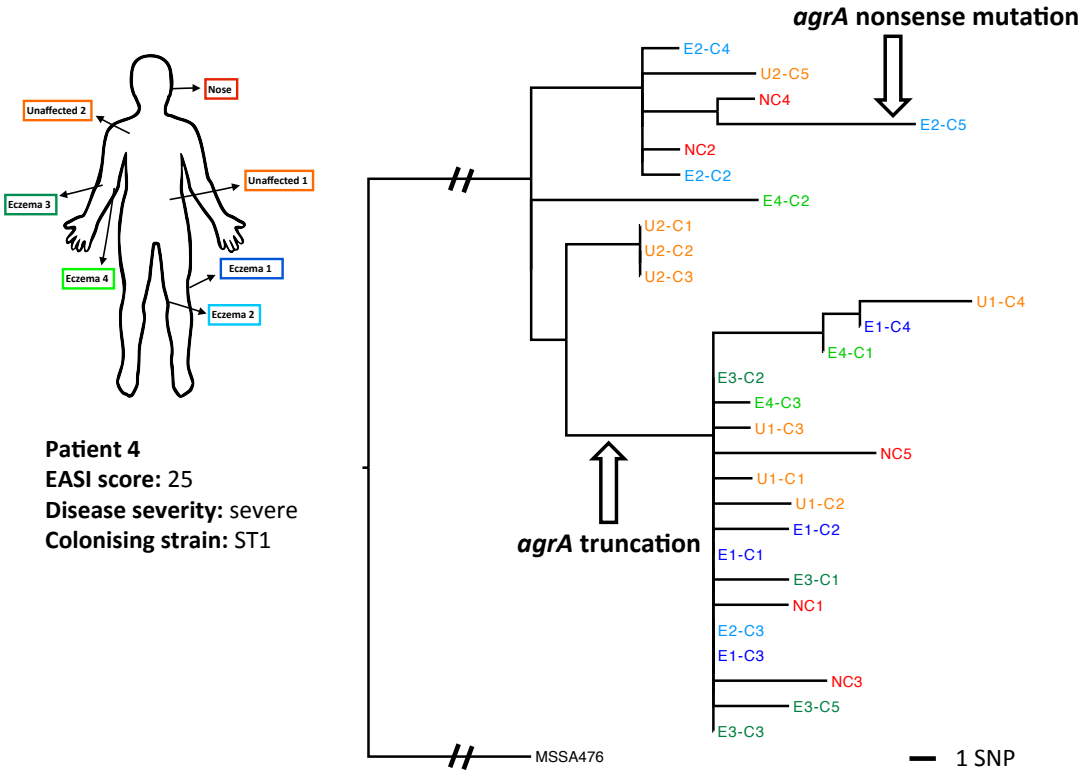

Figure 5-4 Evidence of adaptive mutations in Patient 4 (study ID PSAE004). Maximum likelihood core SNP tree of 28 sequenced colonies from 5 body sites. Branch label marked with black arrows indicate the point in phylogeny where homoplastic mutations in *agrA* (accessory global regulatory system component A) are predicted to have occurred. Tree rooted using MSSA476 reference genome. Labels on body outline diagram indicate sites of sampling, with colouring of branch labels correlating to these body sites. Branch labels: E1/ 3- lateral border of single eczema site. E2/4- medial border of single eczema site. SNP bar is indicated for approximate scale (not applicable to root branches with strikethrough). (EASI- Eczema Area Severity Index).

Table 5-9 Within host heterogeneity identified in Patient 4 (study ID: PSAE004). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in ST1 reference (MSSA476; accession number BX571857). NS- non-synonymous; S- synonymous; I- intergenic; *- premature stop codon.

Patient 4 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
2008647	G->T	NS	SAS1844 putative aldehyde dehydrogenase	A->S	U1-C4, E1-C4, E4-C1, E3-C5, E3-C1, E2-C3, E1-C1, E3-C2, U1-C3, NC5, U1-C2, E4-C3, E1-C2, NC1, U1-C1, E1-C3, NC3, E3-C3
2199110	C->T	NS	SAS2044 hypothetical protein	R->H	U1-C4, E1-C4, E4-C1, E3-C5, E3-C1, E2-C3, E1-C1, E3-C2, U1-C3, NC5, U1-C2, E4-C3, E1-C2, NC1, U1-C1, E1-C3, NC3, E3-C3
836089	G->A	NS	SAS0746 *vacB* putative ribonuclease R	V->I	E4-C2, U2-C1, U2-C3, U2-C2, U1-C4, E1-C4, E4-C1, E3-C5, E3-C1, E2-C3, E1-C1, E3-C2, U1-C3, NC5, U1-C2, E4-C3, E1-C2, NC1, U1-C1, E1-C3, NC3, E3-C3
375374	A->G	NS	SAS0328 putative membrane protein	I->V	E2-C5
573900	C->T	NS	SAS0501 rpoC DNA-directed RNA polymerase beta' chain protein	A->V	E2-C5
2201081	C->T	NS	SAS2046 putative mannose-6-phosphate isomerase	G->S	E2-C5
1834431	C->T	NS	SAS1684 *leuS* leucyl-tRNA synthetase	D->N	E2-C5
577967	T->A	NS	SAS0505 *fusA translation* elongation factor G	S->T	U2-C5
1440802	G->A	NS	SAS1346 putative membrane protein	D->N	NC2
2516189	G->A	NS	SAS2350 putative lipoprotein	D->N	E2-C4
156278	G->A	NS	SAS0135 *capL* capsular polysaccharide synthesis enzyme	A->T	E4-C2
729844	A->G	NS	SAS0651 *mgrA* MarR family regulatory protein	S->P	E4-C2
1137147	G->A	NS	SAS1064 *isdA* iron-regulated heme-iron binding protein	A->V	E4-C2
1308326	T->G	NS	SAS1218 *cinA* competence damage inducible protein	I->M	U2-C1, U2-C3, U2-C2
2738178	G->A	NS	SAS2540 *sraP* putative cell wall-anchored protein	T->I	U1-C4, E1-C4, E4-C1

Patient 4 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
2739483	C->T	NS	SAS2540 *sraP* putative cell wall-anchored protein	S->N	U1-C4, E1-C4, E4-C1
2449841	C->A	NS	SAS2289 tetrapyrrole (corrin/porphyrin) methylase family protein)	A->S	E3-C5
497249	C->T	NS	SAS0438 Orn/Lys/Arg decarboxylase family protein	A->V	U1-C3
683379	G->C	NS	SAS0606 *tagX* putative glycosyl transferase	D->H	NC5
1641468	A->C	NS	SAS1520 *hrcA* heat-inducible transcription repressor	F->V	NC5
461922	C->T	NS	SAS0415a doubtful CDS no known function	A->V	E4-C3
877253	T->A	NS	SAS0794 putative 5'-nucleotidase	I->N	NC1
2308897	G->A	NS	SAS2151 AcrB/AcrD/AcrF family protein	P->L	NC1
2142662	A->T	NS	SAS1985 putative helicase	S->T	U1-C1
837398	G->T	NS	SAS0747 putative tmRNA-binding protein *smpB* SsrA-binding protein	R->L	NC3
2089247	A->T	STOP	SAS1944 *agrA* autoinducer sensor protein response regulator protein	K->*	E2-C5
266915	G->T	STOP	SAS0224 putative zinc-binding dehydrogenase	G->*	U2-C5
685586	A->T	STOP	SAS0608 penicillin-binding protein 4	L->*	E3-C5
304574	G->T	STOP	SAS0258 *esxA* conserved hypothetical protein	E->*	NC5
878792	T->C	S	SAS0796 conserved hypothetical protein DUF1027 superfamily		NC4, E2-C5, U2-C5, E2-C2, NC2, E2-C4
423339	C->T	S	SAS0377 hypothetical protein		E4-C2
1794495	T->C	S	SAS1657 *isdH* haptoglobin-binding surface anchored protein		E4-C2
1461399	C->T	S	SAS1364 conserved hypothetical protein DUF1250 protein		U2-C1, U2-C3, U2-C2
708733	T->C	S	SAS0629 putative phosphate transport protein		U1-C4, E1-C4, E4-C1, E3-C5, E3-C1, E2-C3, E1-C1, E3-C2, U1-C3, NC5, U1-C2, E4-C3, E1-C2, NC1, U1-C1, E1-C3, NC3, E3-C3
1192651	T->C	S	SAS1118 cell division protein FtsQ		E3-C1

Patient 4 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
2771487	G->A	S	SAS2570 putative membrane protein; DUF1393 protein		NC5
1426709	A->G	S	SAS1332 *cvfB* conserved virulence factor B		NC3
1909182	C->T	S	SAS1756 *traP signal* transduction protein		NC3
404091	C->T	I	Between SAS0359 (nitroreductase) and SAS0360 (putative sodium: dicarboxylate symporter protein)		E2-C5
1163710	G->T	I	Between SAS1088 (putative exported protein) and SAS1089 (hypothetical protein)		NC4, E2-C5, U2-C5, E2-C2, NC2, E2-C4
205880	A->G	I	Between SAS0174 (putative permease protein) and SAS0175 (putative membrane protein)		U2-C5
495334	C->T	I	Between rRNA and SAS0438 (Orn/Lys/Arg decarboxylase family protein)		E2-C2
1820894	G->A	I	Between SAS1677 (putative peptidase) and SAS1678 (putative exported protein)		E4-C2
2598270	C->T	I	Between SAS2423 (hypothetical protein) and SAS2424 (PTS system, glucose-specific IIABC component)		U1-C4, E1-C4, E4-C1
1413408	C->T	I	Between SAS1311 (N-(5'phosphoribosyl)anthranilate (PRA) isomerase) and SAS1312 (tryptophan synthase beta chain)		U1-C4, E1-C4, E4-C1, E3-C5, E3-C1, E2-C3, E1-C1, E3-C2, U1-C3, NC5, U1-C2, E4-C3, E1-C2, NC1, U1-C1, E1-C3, NC3, E3-C3
718186	G->A	I	Between SAS0637 (LysR family regulatory protein) and SAS0638 (putative sugar efflux transporter)		E3-C1
2711129	C->A	I	Between SAS2523 (zinc metalloproteinase aureolysin precursor) and SAS2524 (immunodominant antigen B)		U1-C3, NC5, U1-C2
1071847	C->A	I	Between SAS0999 (FolD bifunctional protein and SAS1000 (putative phosphoribosylaminoimidazole carboxylase catalytic subunit)		U1-C2
1035025	T->A	I	Between SAS0963 (hypothetical protein) and SAS0964 (putative exported protein)		E1-C2

Patient 4 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
1035026	A->C	I	Between SAS0963 (hypothetical protein) and SAS0964 (putative exported protein)		E1-C2

Table 5-10 Within host-heterogeneity identified in Patient 4 (study ID: PSAE004). Unique indels differentiating sequenced colonies. Base position and region/ gene are relation to reference assembly used for mapping PSAE004_E3_C5. Region and gene are in position to ST1reference (MSSA476; accession number BX571857). (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
146942	E1C4, E4C1, U1C4	(-) c	SAS0154; putative non-ribosomal peptide synthetase	Truncation of protein
1994907	E2C2, E2C4, E2C5, NC2, NC4, U2C5	(-) t	SAS1760a; hypothetical protein	Doubtful CDS no consequence
2795782	E2C2, E2C4, E2C5, E4C2, NC2, NC4, U2C1, U2C2, U2C3, U2C5	(+) C	SAS1927; hypothetical protein	Frame shift mutation
2811750	U1-C4, E1-C4, E3-C2, E4-C3, U1-C3, NC5, U1-C1, U1-C2, E1-C2, E1-C1, E3-C1, NC1, E2-C3, E1-C3, NC3, E3-C5, E3-C3	(-) t	SAS1944; *agrA* autoinducer sensor protein response regulator protein	Truncation and loss of function
719262	NC3	(+) T	Intergenic between SAS2085 (putative membrane protein) and SAS2086 (glycine betaine transporter 2)	Indel 62bp upstream of start site
441168	E2C5	(-) a	Intergenic between SAS0446 (tetrapyrrole methylase family protein) and SAS0447 (putative methionyl-tRNA synthetase)	Indel 165bp upstream of translational start site

5.7.3.3 Evidence of selection and persistence within the host

Patient 5 was the only case included who suffered with solely localised disease. This two-year-old boy was colonised only at a site of eczema on the dorsum of the left hand. The clinical history indicated a protracted period of disease failing to resolve despite multiple extended courses of topical steroid. Sub-sampling of this site, with two swabs being taken 4 cm apart within the area of affected skin provided the most diverse colony population of all 9 cases. Ten colonies were sequenced from this sampling, in which 26 core genome SNPs (Table 5-11) and 2 indels (Table 5-12) were observed.

What was striking in addition to the maximum SNP distance between the two most genetically distinct colonies of 26, was clear separation of the colony populations dependent on which skin location they were derived from. The two sub-populations were isolated from two swabs taken 4cm apart. As shown in the phylogeny (Figure 5-5) colonies from E1 (eczema 1; medial border of the lesion) are separated as a distinct cluster from E2 (eczema 2; lateral border of the lesion). No intermediary variation is seen outwith these two colony populations, both of which are on long branches, of 11 and 13 SNPs respectively. This suggests the possibilities that either the population has been shaped by a selective sweep removing intervening variants over a period of long-term colonisation or that an intermediary population hasn't been sampled.

In addition to the observed SNPs and indels between these two populations was variance in the MGE content. As described in section 5.6 the colonies varied in their carriage of the plasmid conferred heavy metal resistance determinants. The colonies from the sub-sampling in this child further separated on the basis of variable carriage of a 27 kb plasmid (backbone similar to SAP0194; accession number GQ900385.1) (Table 5-5). Whilst the relevance of

carriage of the plasmid and its encoded determinants is not clear, it is perhaps evidence of the persistence and adaptation of the colonising *S. aureus* to this host.

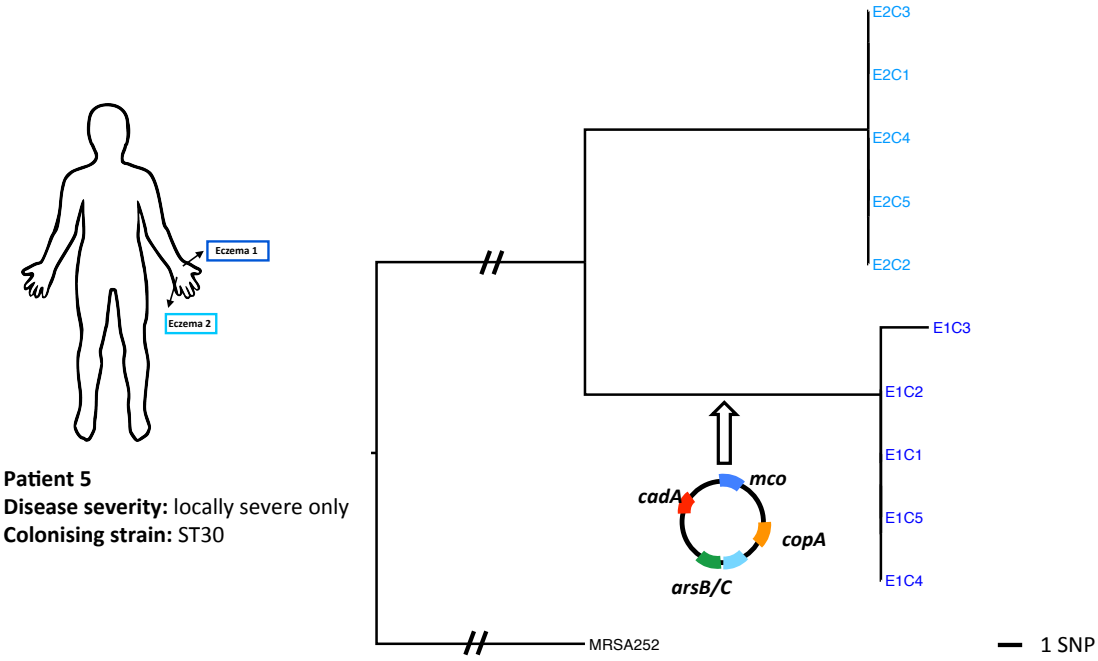

Figure 5-5 Evidence of section persistence within the host in patient 5 (study ID PSAE005). Maximum likelihood core SNPs tree illustrating genetic relationship of 10 colonies from two spatially distinct positions within a single eczema site. Labels on body outline diagram indicate sites of sampling, with colouring of branch labels correlating to these body sites. Branch label marked with a black arrow indicates point in phylogeny where a plasmid carrying metal resistance was gained in the *S. aureus* population. Trees are rooted using MRSA252 reference. Branch labels: E1- medial border of single eczema site; E2- lateral border of single eczema site. SNP bar is indicated for scale (not applicable to root branches with strikethrough). Gene abbreviations on plasmid: *cadA* (cadmium transporting ATPase); *mco* (multicopper oxidase); *copA* (copper exporting P type ATPase); *arsB/C* (arsenical resistance operon).

Table 5-11 Within host heterogeneity identified in Patient 5 (study ID: PSAE005). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in ST36 reference (MRSA252; accession number: BX571856). NS- non-synonymous; S- synonymous; I- intergenic.

Patient 5 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
333222	C->T	NS	SAR0284 *essC* putative membrane protein	H->Y	E2C1, E2-C4, E2-C5, E2-C3, E2-C2
498974	T->A	NS	SAR0471 *gltA* glutamate synthase, large subunit	Y->N	E2C1, E2-C4, E2-C5, E2-C3, E2-C2
776258	G->T	NS	SAR0741 aldo/keto reductase family protein	G->V	E2-C1, E2-C4, E2-C5, E2-C3, E2-C2
941759	C->T	NS	SAR0905 Na+/H + ion transporter family protein	P->L	E2-C1, E2-C4, E2-C5, E2-C3, E2-C2
1423684	G->A	NS	SAR1368 sodium: alanine symporter family protein	G->S	E2-C1, E2-C4, E2-C5, E2-C3, E2-C2
1484337	C->T	NS	SAR1425 *odhA* 2-oxoglutarate dehydrogenase E1 component	M->I	E2-C1, E2-C4, E2-C5, E2-C3, E2-C2
2387452	C->T	NS	SAR2300 *rpsl* 30S ribosomal protein S9	R->H	E2-C1, E2-C4, E2-C5, E2-C3, E2-C2
1117695	G->A	NS	SAR1070 *pdhD* dihydrolipoamide dehydrogenase	A->T	E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
1256390	T->C	NS	SAR1206 *fabD* putative malonyl CoA-acyl carrier protein transacylase	F->S	E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
1333452	A->G	NS	SAR1271 *mutS* DNA mismatch repair protein MutS	R->G	E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
1381855	A->G	NS	SAR1331 *DesK* sensor kinase protein	K->R	E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
1815438	C->T	NS	SAR1752 *hemA* glutamyl-tRNA reductase	A->T	E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
2202764	A->G	NS	SAR2140 *ilvD* putative dihydroxy-acid dehydratase	E->G	E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
2717966	G->A	NS	SAR2632 putative transport protein	T->I	E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
124173	A->G	S	SAR0114 *spA* immunoglobulin G binding protein A precursor		E2-C1, E2-C4, E2-C5, E2-C3, E2-C2
1284617	A->G	S	SAR1230 putative ATP-dependent protease ATP-binding subunit		E2-C1, E2-C4, E2-C5, E2C3, E2-C2
2400014	A->G	S	SAR2320 *rplF* 50S ribosomal protein L6		E2-C1, E2-C4, E2-C5, E2-C3, E2-C2

Patient 5 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
1579426	A->G	S	SAR1485 putative 30S ribosomal protein S1		E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
1643851	T->C	S	SAR1569 *rlu*B ribosomal large subunit pseudouridine synthase B		E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
1683129	T->C	S	SAR1611 putative membrane protein		E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
2228320	C->A	S	SAR2160 putative membrane protein		E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
228302	G->A	S	SAR0198 ABC transporter ATP-binding protein		E1-C3
2254115	T->C	I	Between SAR2183 (*tenA* transcriptional activator) and SAR2183b (hypothetical protein)		E1-C3
2068670	C->T	I	Between SAR1978 (putative metallopeptidase) and SAR1980 (putative membrane protein)		E2-C1, E2-C4, E2-C5, E2-C3, E2-C2
2240567	T->C	I	Between SAR2168 (putative helicase) and SAR2169 (putative UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-alanyl ligase)		E1-C2, E1-C5, E1-C1, E1-C4, E1-C3
2778894	C->G	I	Between SAR2690 (cudB putative choline dehydrogenase) and SAR2691 (cudA putative betaine aldehyde dehydrogenase)		E1-C2, E1-C5, E1-C1, E1-C4, E1-C3

Table 5-12 Within host-heterogeneity identified in Patient 5 (study ID: PSAE005). Unique indels in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to reference assembly used for mapping PSAE005_E1C2. Region and gene are in position to ST36 reference (MRSA252; accession number: BX571856). (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	**Colonies**	**Base change**	**Region/ gene**	**Predicted consequence**
696019	E2C1, E2C2, E2C3, E2C4, E2C5	(-) a	SAR0783 putative membrane protein	Frame shift leading to I->L
853225	E1C1, E1C2, E1C3, E1C4, E1C5	(+) A	At site of plasmid deletion; equivocal to integrated plasmid in MRSA252 region (735184..764371)	Nil

5.7.3.4 Diversity and co-colonisation

Patient 8 provided the only example of co-colonisation amongst the cases, despite deep sampling across the case population. This two and half year girl was carrying both strains (ST45 and ST123) in 3 out of 4 of the colonised sites, including nasally. Both strain populations exhibited differing heterogeneity, with none detected amongst the 6 ST45 colonies. Of the remaining 17 ST123 colonies recovered 4 core genome SNPs separated them (Table 5-13) and no unique indels were found amongst any of the colonies.

The phylogenies of both strains from patient 8 are shown in Figure 5-6. The ST45 colonies are genetically identical but, as also seen in patient 1, there is overlap of colonies derived from across body sites and the nose indicating self-transmission. Similar findings are also evident from the ST123 colonies, with nasal and eczema site colonies being interspersed in the tree. Comparing the distribution of both strains across the patient, the only site positive for *S. aureus*, but not co-colonised was from the left popliteal fossa (labelled in green as E4 in the phylogeny).

The estimated diversities and duration of colonisation in this case are based upon the ST123 colonies (Figure 5-1). The lack of variation amongst the ST45 colonies indicate they represent the most recently acquired clone in this patient, whereas the ST123 colonies have been present for longer (Figure 5-2).

The dual strain colonisation in this case has additional potential clinical relevance. Two clinically important resistance genes differentiate the colonising populations, *fusB* in ST123 and *qacC* in ST45 background (Table 5-5). These both confer reduced susceptibility to examples of the most frequently used therapies in AE, fusidic acid (O'Neill et al. 2004) and benzalkonium chloride (Furi et al. 2013), which is an excipient of topical antiseptic

preparations used in emollients and soap substitutes. This is a further example of heterogeneity in the AMR profile of colonising strains identified in AE cases.

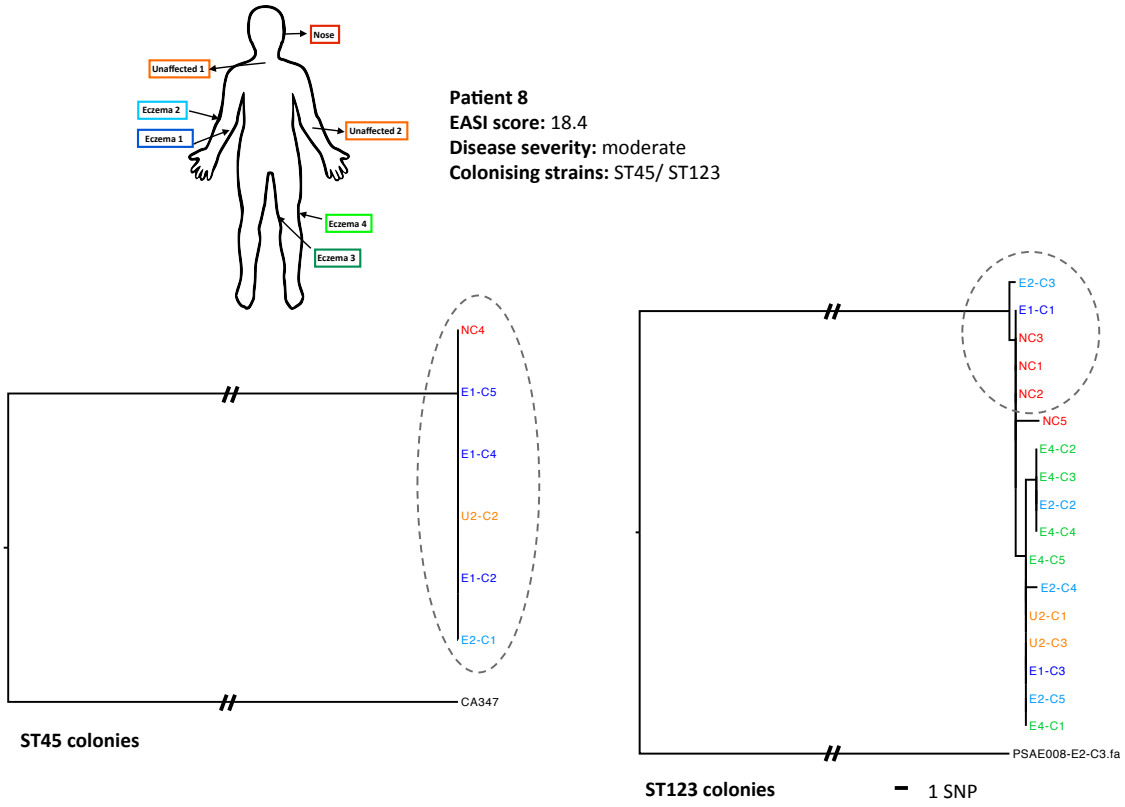

Figure 5-6 Colonisation by two distinct STs in Patient 8 (study ID PSAE008). Maximum likelihood core SNP tree of 23 sequenced colonies from 4 body sites. Both strain types are present nasally and extra-nasally. Branch label colouring corresponds to body site, marked on diagram. Perforated lines indicate overlapping spread between body sites. ST45 colonies are genetically identical. ST45 colony tree routed using CA347 reference; ST123 colony tree routed used self-reference of E2-C3 colony assembly. Branch labels: E1/ 3- lateral border of single eczema site. E2/4- medial border of single eczema site. SNP bar is indicated for scale (not applicable to root branches with strikethrough). (EASI- Eczema Area Severity Index).

Table 5-13 Within host-heterogeneity identified in Patient 8 – ST123 colonies (study ID: PSAE008). Single nucleotide polymorphisms differentiating colonies sequenced colonies. Base position and region/ gene are relation to position in self assembly used for mapping PSAE008_E2C3, with annotation of region transferred from MSSA76 reference genome. NS- non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
1015322	A->G	NS	SAS0090 pyridoxal-phosphate dependent enzyme	S->P	E2-C2, E4-C4, E4-C3, E4-C2, E2-C5, E4-C1, U2-C3, E1-C3, E4-C5, U2-C1, E2-C4, NC5, NC3, NC2, NC1, E1-C1
463142	T->C	NS	SAS2008 *atpA* ATP synthase alpha chain	T->A	NC5
2354961	A->T	NS	SAS1966 *ilvA* threonine dehydratase biosynthetic	I->N	NC5
1664972	T->C	NS	SAS0668 aldo/keto reductase family protein	K->R	E2-C2, E4-C4, E4-C3, E4-C2

5.8 Comparison of colonisation dynamics and heterogeneity between cases and controls

Physiologically, the inflamed skin of AE represents a distinct colonisation environment in comparison with nasal epithelium, which is a niche habitat for *S. aureus*. To gain further understanding of the colonisation of AE skin by *S. aureus,* this study population was compared with the healthy nasal carriage group described in chapter 4. This comparison aimed to look for evidence of differential diversification and environmental selection arising on these differing cutaneous sites.

5.8.1 Diversity and duration of colonisation in cases and controls

Firstly to assess if the dynamic process of colonisation varied according to its site, i.e. healthy nasal epithelium vs. inflamed extra-nasal epithelium, the observed diversities (SNPs per colony) from each group were compared.

As shown in Figure 5-7 cases and controls exhibited similar range of diversities across their respective study groups. Nasal controls and cases are interspersed throughout this gradient, with examples of low-level diversity being seen in cases such as PSAE002 and 8 as well as controls SS_099 to 147. Equally cases and controls are also present at the most diverse end of the spectrum. The greatest diversity found in the studies is within nasal controls SS_149 and 105, but this is followed closely by case PSAE005. These results indicate that the cutaneous colonisation site did not influence the accumulation of within host diversity.

The genetic diversity of the clonal *S. aureus* populations is suggestive of differing periods of colonisation in the individuals. Whilst it cannot be ruled out that some of these populations arose in the individual via multiple transmissions from a clonal source, the clonal expansion is consistent with within host diversification. Predicted short durations of colonisation at the time of sampling can be seen in in both cases and controls. Both study groups are represented through this range of less than 1 week to 9 months. The individuals estimated to be colonised for the greatest longest periods are not necessarily nasal carriers. This was perhaps surprising. Asymptomatic carriage of this organism can persist over months to years (Kluytmans et al. 1997; Wertheim et al. 2005), although no definition of carriage status can be applied to these controls as sampling took place at just one time point. It is interesting to note that multiple cases show relatively old estimated colonisation populations, especially given that these are body sites which have been continually exposed to bathing and topical therapies for their AE.

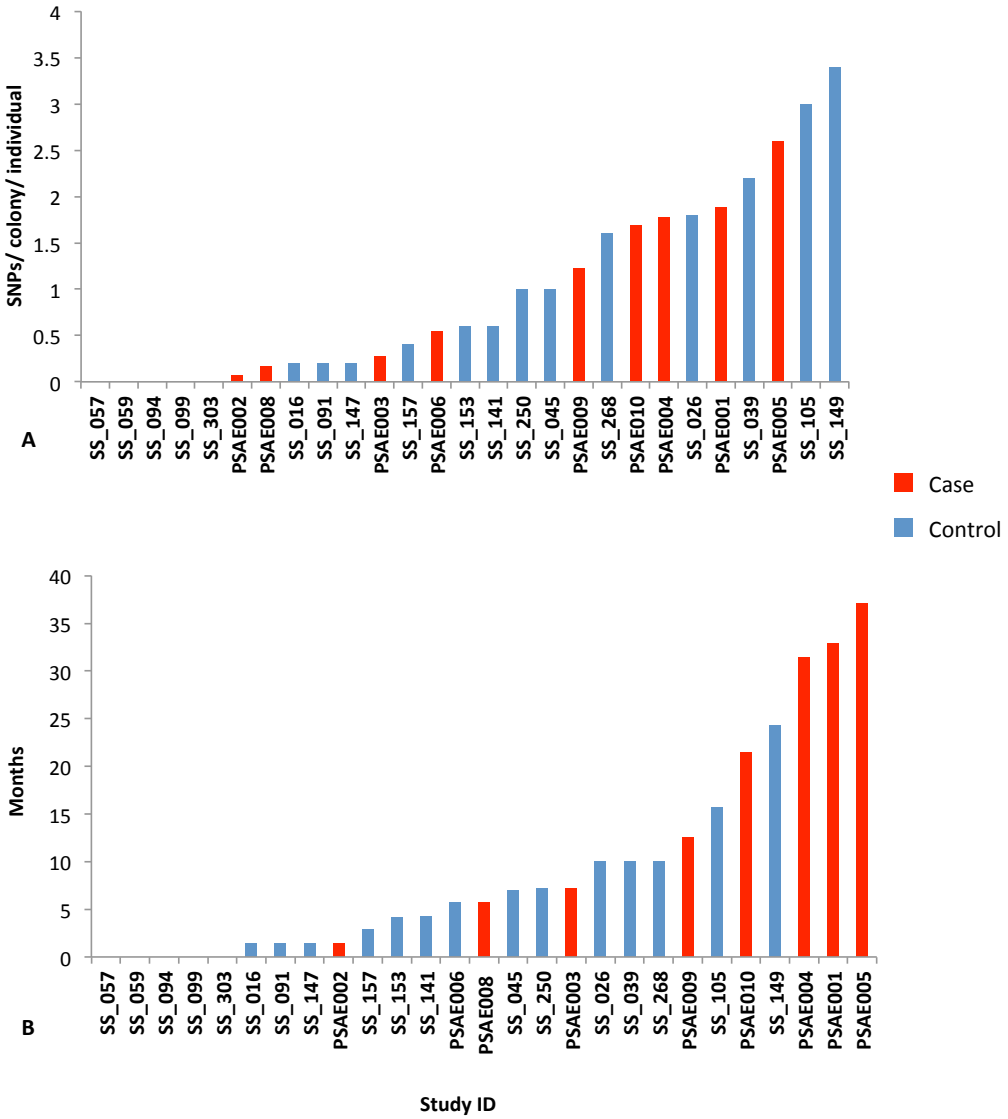

Figure 5-7 Diversity and estimated age of the colonisating population in cases and controls. A- comparison of diversity (SNPs per sequenced colony per individual) in colonisation in healthy nasal carriers and AE cases. **B**- comparison of the estimated age of the colonising population (months) in nasal carriers compared with AE cases.

5.8.2 Differentiation of *S. aureus* during colonisation: case control comparison

Diseased skin represents a very different environment physiologically in contrast to the nasal epithelium. It is comparatively much more exposed, and subject to environmental influences such as differing temperature, bathing and topical treatments. To look for evidence of differential diversification and selection on the basis of environmental influences, the frequency SNP types in cases and controls were compared. For both study populations the distribution for each class of base substitution was compared.

A comparison of the proportions of non-synonymous, synonymous and intergenic SNPs between case and control isolates was made to assess if there was evidence of differential selection within the hosts on the basis of disease site or nasal colonisation. Non-synonymous mutations accounted for 51.6 % (n= 46) of SNPs in controls compared with 54% (n= 106) of SNPs in cases. Synonymous mutations made up 20.2% (n=18) of SNPs in controls and 22.4 % (n= 44) in cases. Finally intergenic SNPs made up 28% (n=25) of SNPs in controls and 23.9% (n=47) in cases. This revealed that there was no significant difference in the overall proportion of these different SNP types occurring between the case and control study groups (Figure 5-8) ($p=0.74$; Chi2, two-tailed test).

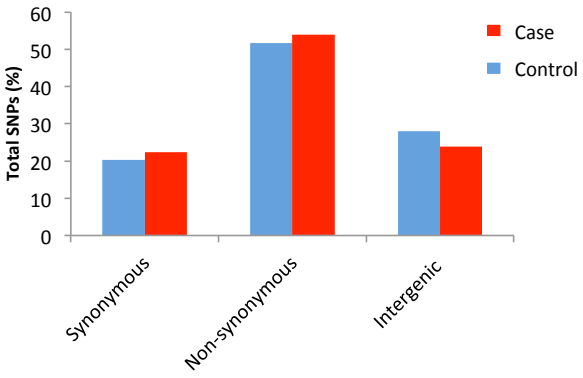

Figure 5-8 Comparison of SNP types between cases and controls. Total number of core genome SNPs in cases and controls= 286 (Total number of all core genome SNP across 9 cases = 197; Total number of SNPs found in 18 controls = 89).

On the premise that differentiation arising within the host may be due to the influence of the different colonisation environment posed by the anterior nares or inflamed extra-nasal skin, may not be evident from comparison of the SNP types, comparison of the functional classes of genes accumulating mutations were compared between the groups. All non-synonymous, and hence potentially functionally relevant SNPs, across cases and controls were assigned to a gene class on the basis of inferred or previously defined function using a modified version of Riley classification scheme (Riley and Space 1996).

In total 152 non-synonymous core SNPs were used for this comparison across the cases and controls. These were distributed across 23 functional gene classes across the *S. aureus* genome with several categories seemingly differentiating the study groups (Figure 5-9). In the controls, membrane, transport, chaperone, protective response and central intermediary metabolism protein classes appeared to be enriched. In cases, hypothetical proteins, sensor kinases, Gram-positive surface anchored, and proteins associated with macromolecule degradation were more frequently mutated. Comparison of the number of non-synonymous SNPs in the various functional categories different between cases and controls were not statistically significantly

different (p=0.27; Fisher's exact test, two tailed). Limited sampling in this study evidently provides a short timescale to observe diversification during in host clonal expansion. Therefore it is unlikely that the sampling was saturated enough to generate clear signals of selection on the genomic scale. Consequently there is insufficient power to statistically differentiate cases and controls due to small number of SNPs and limited coverage of the genome provided by this dataset.

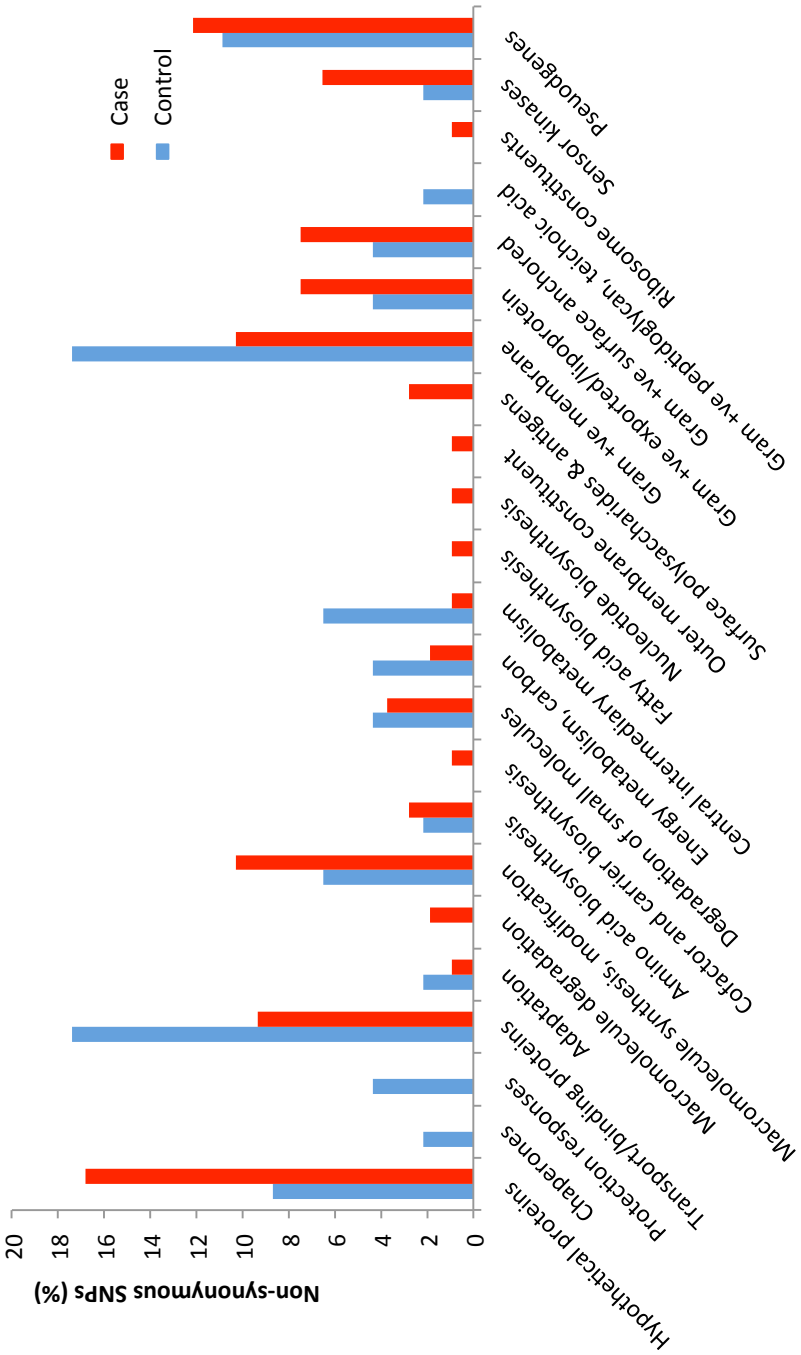

Figure 5-9 Distribution of non-synonymous base substitutions across the *S. aureus* genome by functional gene classes. Percentage of non-synonymous SNPs identified within gene class in cases and controls.

5.9 Discussion

5.9.1 Diversity of *S. aureus* in atopic eczema disease flares

Multiple studies over decades demonstrated a link between increased disease activity in atopic eczema and colonisation by *S. aureus* (Williams et al. 1990; Goh et al. 1997; Guzik et al. 2005; Tauber et al. 2016). In spite of this little is currently understood about this organism as a normal component of the human microbiota and how it contributes to AE disease aetiology. Using WGS to deep sequence *S. aureus* colonising children at the time of flares in their AE has allowed an unparalleled view of the organism in children with active disease. It has allowed insights to be gained on the clonality of colonising population, transmission, duration of carriage and how clinical interventions may impact upon the *S. aureus* populations.

Clonal expansion in colonisation

A recent metagenomic study clearly demonstrated that during increasing disease severity in eczema there is expansion of the *S. aureus* population which occurs at the overall expense of cutaneous microbial diversity (Kong et al. 2012). The in-depth examination of colonising populations from children with active AE in this work has demonstrated that the increase of the *S. aureus* population described by Kong *et al* (2012). is the direct result of a clonal expansion within host. One of the initial aims of this study was to determine if there was strain heterogeneity associated with disease flares in eczema. Previous studies in eczema have almost universally relied upon single colony assessment, and generally from single body sites (Lomholt et al. 2005; Rojo et al. 2014) meaning that this has never before been addressed. Co-colonisation was identified in patient 8 and within 3 sampled body sites, which validated the approach of multiple colony assessment.

Evidence of self-transmission

The primary reservoir of *S. aureus* colonisation in eczema is frequently highlighted as an unanswered question in this field of research. Several authors have indicated the nose as the likely primary reservoir (Hoeger et al. 1992; Lomholt et al. 2005), but again the majority of studies do not take this into consideration when screening children with AE for *S. aureus* colonisation. From the results there are clear examples of self-transmission, with overlap of nasally derived colonies and extra-nasal body sites (Patient 1/ 4/ 8). While it cannot be definitely concluded that nasal colonisation in these children was present prior to extra-nasal carriage, the basal position of nasal colonies in the phylogenies in patients 1, 4 and 8 (Figures 5-3/4/6) would suggest this to be the case. Five of the AE cases, however, were not found to be nasal carriers at the time of sampling. This raises the possibilities of either transitory nasal carriage or their extra-nasal colonisation deriving from an extrinsic source. In these children, transmission could be the result of close contact with a carrier such as a parent or from their home environment. The alternative possibility is that in these individuals there were self-reservoirs of *S. aureus* that were not sampled. In addition to the nares as a niche colonisation site, other sites such as the groin and throat are other potential reservoirs (Wertheim et al. 2005; Esposito et al. 2014), which could have led to self-transmission. Longitudinal follow up and sampling of affected individuals and their family would be needed to understand this further.

Diversification during carriage

Diversification of a colonising population in AE has never before been demonstrated although it has been well described in nasal carriage and transmission (Golubchik et al. 2013; Harris et al. 2013; Paterson et al. 2015). The association between disease severity and increasing *S. aureus* burden has been previously shown (Kong et al. 2012), but the temporal latency for AE-associated colonisation has not been reported. The methodology used in this study has allowed this to be inferred. Based upon the temporal calibration derived from the diversity of

the sampled cases, several were estimated to contain populations that have diversified from a common ancestor over longer time periods than others. Strikingly the levels of diversity in some cases was comparable with that observed in nasal colonisation of greater than 6 months duration (Young et al. 2012).

Attenuation of virulence during colonisation

In cases where these mature colonising populations were uncovered there was evidence of selection having shaped the populations. These examples have provided insights into the mechanisms used by this organism to persist in host. The primary example of this is patient 4, where the identified homoplastic mutation of *agrA* (Figure 5-4; Tables 5-9/10) are strongly indicative of selection in host. Survival of *S. aureus* intracellularly in keratinocytes is well recognised (Eiff et al. 2001; Sendi and Proctor 2009) and in AE this may be a source of persistent and recurrent infection. A recent study demonstrated that strains bearing *agr* system mutations were able to internalise in keratinocytes and persist within the cells (Soong et al. 2015). The authors also reported that 22% of isolates studied from individuals with recurring *S. aureus* infection complicating their atopic eczema had this *agr*- mutant phenotype. Therefore, in patient 4 it is possible that intracellular survival has contributed to the persistent carriage. This finding highlights a potential in host adaptation supporting colonisation that in eczema could lead to a persisting *S. aureus* population and therapeutic failure.

Accessory genome variation in-host

The instances of accessory genome variability were also found in cases where colonisation was estimated to have been by more mature populations. In patient 5 there was a variably present plasmid, although not carrying obviously clinically relevant resistance determinants, it does highlight population heterogeneity within a single sampling. In patient 1, the plasmid that varied in the colonising population carrying both *blaZ*, and a bacteriocin (anti-microbial peptide) and immunity protein, which may be of biological relevance. Prokaryotes use

bacteriocins to inhibit growth of and kill closely related organisms (Heng et al. 2007). In this patient colonies carrying the plasmid in this case were predominantly within the extra-nasal population where the potential to inhibit other microbial biota by this means could have beneficial. Additionally, in these cases the phylogeny show examples of deep branches separating the colonising populations into two sub-populations, which suggestive of a selective sweep having removing intermediate genetic variation from the *S. aureus* population. Both examples again serve to emphasise evidence of diversification over a protracted course and phenotypic heterogeneity that may impact on their ability to survive and thrive despite competition and clinical interventions.

Antibiotic resistance and its potential consequences

Antimicrobial interventions are a frequent occurrence in children with eczema. Evident from the sequencing data is the impact that this has on individuals colonising populations and the potential implications for the disease-promoting capacity of the strain. The identification of the *fusA* mutation in patient 1 (Table 5-5; *fusA* L461K), for instance, is evidence of a core genome mutation conferring resistance to the common therapy, fusidic acid. This in the context of a clinical history confirming a course of topical fusidic acid several months prior to sampling. The *fusA* gene encodes translation elongation factor G (EF-G), a GTPase catalysing the elongation phase of bacterial protein synthesis on the ribosome (Hansson et al. 2005). Fusidic acid therapy works by inhibiting bacterial protein elongation, and mutations in *fusA* confer resistance by changing the affinity of elongation factor G for the ribosome, as well as fusidic acids affinity for EF-G (Hansson et al. 2005). The *fusA* L461K mutation has an association with a small colony variant (SCV) phenotype in *S. aureus* (Norström et al. 2007). Small colony variants are phenotypic sub-populations which are both slower growing and show increased propensity to persist intracellularly in the host (Sendi and Proctor 2009). Relapsing infection and treatment failure have been reported due to SCVs in numerous contexts including cystic

fibrosis (Kahl et al. 1998), chronic rhinositis (Hayes et al. 2015) and of more relevance the genetic skin disorder, Darier's disease (Eiff et al. 2001).

5.9.2 Diversity in *S. aureus* colonisation between nasal carriage controls and AE cases

The comparison between nasal carriage controls and AE cases allowed an exploration of the colonisation dynamics of *S. aureus* across nasal and diseased extra-nasal cutaneous environments. Whilst colonisation heterogeneity has been characterised in adults and primarily in hospital based populations these studies focussed on community and outpatient based paediatric populations. This has allowed characterisation of *S. aureus* carriage in an otherwise healthy population of children, outwith disease states and therapeutic interventions which are more frequent in the adult population and may influence *S. aureus* carriage.

Clonal colonisation in nasal carriage and AE

This case-control comparison has found the same clonal population structure amongst *S. aureus* isolates from healthy paediatric nasal carriers and children affected by eczema. This demonstrates that observed population sometimes within the AE cases is not an influence of the underlying disease. This in itself is an important observation. Atopic eczema is a complex disorder of epidermal barrier dysfunction and altered cutaneous immunity (Irvine et al. 2011; Leung and Guttman-Yassky 2014). Although the intricacies of this disease are not touched upon in this study, it does not appear from these findings that atopic skin inflammation alters the natural colonisation dynamic of this organism on human skin in comparison to what was observed in nasal carriage controls.

Within host diversity during asymptomatic carriage and in AE

Our observed levels of diversity in cases and controls were similar to previous reports (Young et al. 2012; Golubchik et al. 2013). This suggests that carriage on inflamed skin as opposed to intact nasal epithelium has not influenced the diversity observed. However it must be stressed

that the small sample size and single time point sampling of these studies are limitations that mean this cannot be firmly concluded. If a larger sampling study was undertaken, it could be hypothesised that there would be evidence of differences in the diversity observed for several potential reasons. For instance, the exposed nature of this skin in AE may make it more prone to repeated colonisation events where another hosts' diversity could be transmitted, but also the potential impact of a wide range of topical therapies might drive genetic diversification. Interestingly, whilst there were examples of estimated long-term carriage, and hence diverse populations in the healthy controls, there were no examples of accessory genome variation. However, this assessment did not extend to the depth reported by Paterson *et al* (2015) where individuals' samples were assessed for variable phage carriage. This study was primarily directed at looking for clinically relevant AMR determinants, which in turn uncovered accessory variation in plasmid carriage in two AE cases (patient 1 and 5). Why this was only observed within cases is not known, but it is interesting to consider that in this case there may be an environmental influence of eczema-affected skin on the MGE carriage. For instance, the overall microbial composition of extra-nasal skin is known to differ from the anterior nares (Grice et al. 2009; Grice and Segre 2011), and these sites are also continually exposed to eczema therapies for the management of eczema, which impact upon the microbiome (Kong et al. 2012). These factors therefore may have influenced the findings in AE cases.

Differentiation of S. aureus during colonisation of differential cutaneous environments

Whilst comparison of functional gene classes affected by mutations across the clinical groups did not reach statistical significance, the single time point sampling in this study gives a narrow window through which to observe adaptation in the host. There was, however, a plausible signal of differentiation in the contrasting cutaneous environments being studied. For instances sensor kinases stand out as being biologically relevant in eczema given their role in environmental sensing and virulence regulation (Krell et al. 2010). There was an accumulation of mutations in this class both within and between cases, as seen in patients 9 and 4. In patient

4 three sensor kinases had accumulated non-synonymous mutations including: *smpB* (tmRNA-binding protein), *mgrA* (MarR family regulatory protein), and *hrcA* (heat-inducible transcription repressor) (Table 5-9) and there were two separate loss of function mutations in a fourth sensor kinase, *agrA* (Table 5-9/ 10). In patient 9, SNPs were found in a putative transcriptional regulator of the AraC family and *resE,* a component of the Staphylococcal respiratory response regulatory (Srr) system (Supplementary Appendix B). Within controls, a single sensor kinase gene SNP was found, *malR* (maltose operon transcriptional repressor) in SS_250 (Supplementary Appendix A) and a single nonsense mutation in *agrA* in SS_105 (Table 4-7-1, Chapter 4). Several of these genes are known to have key functions in virulence regulation, for instance an example of an AraC transcriptional regulator mutation identified in a long-term *S. aureus* carrier has recently been shown to have profound consequences on virulence and capacity to cause invasive disease (Young et al. 2012; Das et al. 2016). The *mgrA* and Ssr system genes are specifically involved in virulence regulation under conditions of environmental metabolic stress (Yarwood et al. 2001; Somerville and Proctor 2009). These results although only representative of single instances indicate a pattern of accumulation of mutations that would reduce the overall virulence of the colonising strain occurring during carriage, in populations with estimated ages of 7 months or more.

Observation of pathogens during carriage and colonisation are powerful routes to understanding adaptation that supports survival in the host, which can be used to explore their contribution to the disease pathogenesis. This has been shown in the context of *Pseudomonas* colonising children with cystic fibrosis, where convergent evolution was shown to shape the regulatory networks involved in biofilm formation and virulence (Marvig et al. 2015). These authors proposed that some of the observed mutations arising in the *Pseudomonas* carriage strains represented a shift towards the chronic infection state, with increased production of biofilm components for instance. In the context of *S. aureus* carriage the *agr* system mutations are potentially an example of this. Mutations in this system in

addition to reduced virulence have been shown to be an adaption to subvert the host immune system and establish intracellular populations (Tuchscherr et al. 2011; Soong et al. 2015). Their prevalence in clinical isolates of 15% (Shopsin et al. 2008; Traber et al. 2008) would suggest that this is potentially one of the most frequently adopted mechanisms to survive on a long-term basis in the host. The possible relevance of the other observed mutations found between cases and controls is not clear, but longitudinal assessment of carriage in these groups would undoubtedly be the way forward to provide further clarity.

5.10 Concluding remarks

5.10.1 Diversity of S. aureus colonising children with AE

This study has demonstrated the diversification of *S. aureus* during colonisation of children with AE. The apparent diversity of carriage in some cases suggest that intracellular populations may be an under-recognised phenotype in eczema, contributing to recurring disease flares and failure of response to therapy. The importance of accurate identification of the source of colonisation to direct rationalised therapy is also suggested from this analysis. In addition to driving antimicrobial resistance these interventions may alter the disease-promoting capacity of the colonising strain. Longitudinal follow up of colonisation in children with AE to assess diversity, evidence of adaptation and the consequences of therapies will allow advance of understanding on how *S. aureus* contributes to disease activity in AE, and particularly in the sub-group of children prone to repeated infective exacerbations.

5.10.2 Diversity of *S. aureus* carriage in healthy children versus AE cases

Comparison of colonisation in these two paediatric populations has clearly demonstrated of the similarity of colonisation dynamics occurring in nasal carriers and AE cases. Clonally expanded populations colonise both study groups, both nasally and on inflamed extra-nasal skin. Similar levels of diversity and durations of carriage between the groups indicate that these are also a reflection *S. aureus* colonisation in general outwith a disease context. Whilst functional gene class assessment in the studies failed to differentiate the populations due to the small sample size there were signals suggesting divergence on the basis of disease status. This again illustrates the importance of longer-term observation of colonisation as a means to understanding the potential differential adaptation of the organism to the different host epithelial environments.

6 *Staphylococcus aureus* associated with Atopic Eczema: case control study

6.1 Introduction

The works presented in Chapters 4 and 5 used in depth sampling and sequencing to characterise the colonising populations of AE cases and controls. In direct contrast to this micro-epidemiological approach the sample collection used for this chapter provided the opportunity to study the relationship of *S. aureus* with AE disease from a macro-epidemiological perspective. Multiple studies have previously assessed disease-associated *S. aureus* isolates in AE, looking for evidence of strain association (Kim et al. 2009), whether such strains are more toxigenic (Schlievert et al. 2008), as well as whether multiple strain colonisation is a factor (Lomholt et al. 2005). Crucially, these works have had to rely upon limited genetic typing methods such as PFGE and spa typing for strain characterisation and PCR detection of target genes has meant a limited breadth of assessment was possible. This sample collection therefore provided the opportunity to assess disease and carriage isolates to a greater depth using WGS.

The sample collection utilised for this study was collected and kindly provided by our collaborators Professor Alan Irvine, Dr Maeve McAleer (National Children's Research Centre, Our Lady's Children's Hospital, Dublin) and Dr Desiree Bennett (Epidemiology and Molecular Biology Unit, Temple Street Children's University Hospital, Dublin).

The DNA library preparation and whole genome sequencing for this study was kindly carried out by Dr Kerry Pettigrew, Infection Group, University of St Andrews.

6.2 Aims and objectives

The aim of this study was to assess the genetic characteristic strains of *S. aureus* associated with atopic eczema disease flare in contrast with those found in asymptomatic nasal carriers. Comparative genomic analysis was used to compare the population structure of this isolate collection and examine variation in gene content of isolates between disease states. Previous

work in this field has tended to be virulence determinant-centric. This study therefore aimed to take a more holistic approach assessing the population structure with regards to strain background, and then assessing content of genetic determinants that may confer the ability to colonise and cause disease inclusive of toxins, resistance genes and adhesins.

This work was undertaken as an observational study to apply a breadth of assessment that prior work in this field has not able to extend to given their utilised methods. From the outset it was accepted that the primary limitation of this study would be sample size, which is due in part to the difficulty of obtaining samples from treatment naïve individuals with AE. The sample size for this study was based upon the material available for the study, and is similar to previous studies of this type. A retrospective power calculation is not considered good scientific practice and was therefore not undertaken. When appropriate confidence intervals, which take into account sample sizes, were used to assist with interpretation of findings.

6.3 Recruitment criteria and sampling

AE cases

Atopic eczema cases (age 0 to 7 years) recruited for this study presented to the tertiary referral service at Our Lady's Children's Hospital, Crumlin, Dublin, Ireland between September 2012 and September 2014. They were assessed and diagnosed as having AE by an experienced paediatric dermatologist and with disease severity scoring by SCORAD (SCORing AD) (Oranje et al. 2007) in keeping with recommendations for reporting of AE disease in clinical studies (Schmitt et al. 2014).

The following exclusion criteria were applied during recruitment of cases: pyrexial illness in the preceding 2 weeks, any immunosuppressive therapy (including systemic steroids) in preceding 3 months, or oral antibiotics within previous 4 weeks. Recruited cases were then swabbed

from the single most active eczema lesion at time of assessment, which was clinically presumed to be colonised by *S. aureus*.

Nasal carriage controls

Nasal carriage study participants (age 0 to 7 years) were recruited during attendance to the Emergency Department of Temple Street Children's University Hospital during July and August 2009. Study recruits were assessed and invited to take part if their presentation was not related to infection. Exclusion criteria for this study included: pyrexia, diagnosis of infection at time of presentation, history of atopic disease (including eczema/ hayfever/ asthma/ food allergy), or history of inflammatory skin disease. Recruited controls were then sampled by swabbing a single nostril.

6.4 Clinical phenotype of recruits

A total of 99 *S. aureus* isolates were used for this study. These were derived from clinical swabs taken from children with atopic eczema, or from children attending A&E for a non-infectious illness. In contrast to the works presented in chapters 4 and 5 a single representative colony from the swab was used for this study.

The clinical phenotype of the participants is presented in Table 6-1. Disease severity scoring in this study was undertaken using the SCORAD system, which assesses the same disease features as the EASI score described in Chapter 5, but additionally takes into account scoring of parents reporting of their child's symptoms.

Table 6-1 Atopic eczema case and control characteristics. SCORAD disease severity grading; mild 0-24, moderate 25-50, severe 51-103 (Oranje et al. 2007).

Phenotype	Atopic eczema case	Nasal carriage control
Total number of participants	50	49
Age range (years)	0 to 7	0 to 5
Sex	31 M, 16 F (3 unknown)	22 M, 26 F
SCORAD (range)	7.4 to 91.3	N/A
Atopic eczema	50	0
Other atopic disease	Incomplete data	0

6.5 Clonal complex breakdown between cases and controls

All 99 single colony isolates were whole genome sequenced including 50 cases and 49 nasal controls. From the assembled genomes the MLST profile was extracted and isolates were then grouped into their respective clonal complexes in order to compare the genetic make up of the two disease groups (Figure 6-1). This revealed diverse populations of isolates occurring across cases and controls. Among cases, a total of 18 different sequence types were identified, belonging to 10 clonal complexes. Among controls, 20 different STs were found deriving from 9 clonal complexes.

The overall clonal structure of the isolate collection is shown in Figure 6-1. As demonstrated in the previous chapters there is evidence of differentiation between cases and controls with respects to the CCs identified colonising these children. Clonal complex 1 isolates were the most predominant genetic background in cases accounting for 20% (n=10) of the strains, in comparison to 8% (n=4) of controls although this difference of 12% (95% CI -2% to 25%; Table 6-2) was not detected as statistically significant, at the threshold of significance chosen, in this small study (p= 0.091; Table 6-2). Conversely a significant difference in carriage of CC30 isolates were found in 33% (n=16) of nasal carriers compared with 10% (n=5) of AE cases (p= 0.0058; Table 6-2). Isolates from CC 7, 9 and 59 isolates were present only in cases, each

representing 6% (n=3) of the total of isolates in AE disease cohort but again this difference was not significant for any of these CCs (p= 0.082; Table 6-2). Isolates from CC22 were found only in nasal controls (10%; n=5) (p= 0.02; Table 6-2) and as were CC25 isolates, (2%; n=) (p= 0.31; Table 6-2). CC45 were also more frequent in controls (22.5%; n=11) compared with cases (14%; n=7) (p= 0.28; Table 6-2).

Table 6-2 Chi2 comparison of colonising strain backgrounds in cases and controls. * indicates p <0.05, ** indicates p value <0.01.

Clonal complex or ST of colonising strain	Cases (%)	Controls (%)	Difference (%)	95% Confidence interval for difference	P- value
CC1	20	8	12	-2% to 25%	0.091
CC5	12	16	4	-18% to 9	0.54
CC7	6	0	6	-1% to 13%	0.082
CC8	14	2	12	2 to 22%	0.029**
CC9	6	0	6	-1% to 13%	0.082
CC15	6	2	4	-4 to 12%	0.32
CC22	0	10	10	2 to 19%	0.02**
CC25	0	2	2	-2% to 6%	0.31
CC30	10	33	23	7 to 38%	0.0058**
CC45	14	22	8	-7% to 24%	0.28
CC59	6	0	6	-1% to 13%	0.082
CC121	2	2	0	-6% to 6%	0.9
ST779	2	2	0	-6% to 6%	0.9
ST1290	2	0	2	-2% to 6%	0.32

The high prevalence of CC30 and 45 isolates in the nasal controls would be in keeping with the findings of carriage studies in the European population (Melles et al. 2004; Monecke et al. 2009). Clonal complex 1 isolates are comparatively less common amongst nasal carriage studies, which makes their presence within cases at more than twice the frequency of controls notable.

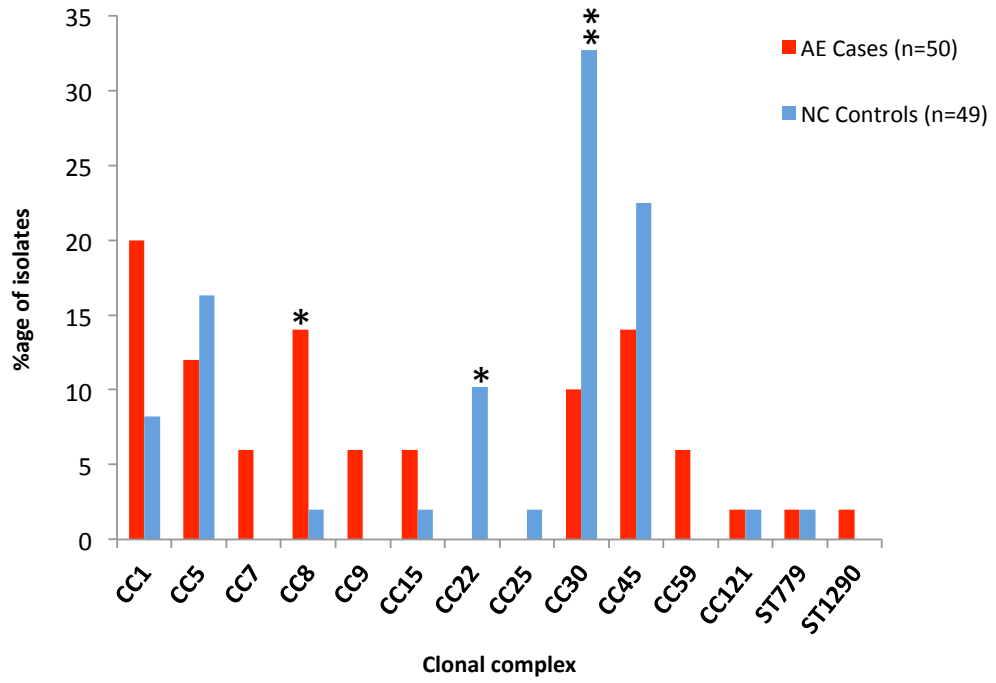

Figure 6-1 Clonal complex breakdown of isolates from age-matched AE cases and nasal carriage controls. Singleton isolates falling outwith defined clonal complex (CC) are denoted by their sequence type (ST). * indicates p <0.05, ** indicates p value <0.01 (as per Table 6-2).

6.6 Comparison of virulence gene content between disease and carriage isolates

Previous studies assessing AE disease-associated *S. aureus* isolates have taken a target gene approach, focussing on the presence or absence of genes hypothesised to contribute to the disease aetiology. Using WGS in this study has allowed a more in depth assessment of the gene content between disease and carriage.

Give the apparent segregation of disease and carriage populations on the basis of the genetic background of the isolates, the gene content across 3 broad classes of *S. aureus* virulence determinants including immune-modulatory, resistance and adherence, were compared between the groups. Virulence factors were identified through review of the literature relating to *S. aureus* in AE, as well as *S. aureus* in general to provide a more comprehensive assessment than previously published in the AE field.

6.6.1 Toxins and immune evasion determinants

One of the defining features of AE as a disease is T cell mediated cutaneous inflammation. The ability of certain staphylococcal toxins to stimulate T cell activation has made them a particular focus of research in this field. Typically enterotoxins A, B, C and TSST are the most frequently studied (Bunikowski et al. 2000; Zollner et al. 2000; Schlievert et al. 2008). The production of SEB (enterotoxin B) for instance, has been shown to correlate with disease severity, which mechanistically in AE is believed to be driven by T cell activation and production of toxin specific IgE (Zollner et al. 2000; Ardern-Jones et al. 2007). Other studies have adopted the alternate approach of characterising gene content by PCR detection or microarray and making an association with disease on the basis of comparison between disease groups (Lomholt et al. 2005; Kim et al. 2009; Yeung et al. 2011; Rojo et al. 2014).

The virulence gene content of the isolates was compared between cases and controls to determine if there was evidence of a difference on the basis of disease status. An allelic database of potential toxin and immunomodulatory genes was compiled from the literature, and all isolates were then mapped to this using SRST2 (Inouye et al. 2014), to identify their presence or absence (Chapter 2: Materials and methods; Table 2-13). A set of 30 virulence genes that had previously been linked with causality in eczema, or belonging to the same gene class were utilised for this comparison. This included the haemolysins and phenol soluble modulins of which delta toxin has been shown functionally to contribute to AE disease (Nakamura et al. 2013). Also investigated was the immune evasion complex locus including staphylococcal complement inhibitor (*scin*), chemotaxis inhibitory protein (CHIPS), and staphylokinase (*sak*), of which *scin* has been associated with disease severity (Rojo et al. 2014). The enterotoxins A to L were compared given previous link to disease (Zollner et al. 2000), the superantigen like protein toxin locus given potential similarity to the enterotoxins, and several toxins associated with skin disease in the wider context such as exfoliatin A, were also included. The initial comparison made was prevalence of these genes across disease and

carriage populations as has also been undertaken in previous similar studies (Kim et al. 2009; Rojo et al. 2014).

The comparison presented in Figure 6-2 suggested a difference in the virulence determinant content of the isolates between disease groups. The haemolysins alpha and delta, as well as phenol soluble modulin β1 were identified in all 99 isolates. Also present at the same frequency in cases and controls were enterotoxins D/J and exfoliatin A. More prevalent within cases were enterotoxins A/B/H, leukocidins D and E, psmβ2 and superantigen like proteins 2 to 11 (Table 6-3). Nasal carriage controls were more frequently carrying staphylokinase (sak), chemotaxis inhibitory protein, enterotoxins C/G and L, toxic shock toxin (tst) and superantigen like proteins 1 and 2 (Table 6-3). No instances of the Paton-Valentine Leuckocidins *lukS* or *lukF* (PVL) were identified.

These results suggested differences between disease and carriage groups in clinically relevant genes. Enterotoxins A and B, for instance were more prevalent in cases, and have both previously been associated with clinical disease severity and with immunomodulation *in vitro* by demonstration of T cell activation (Bunikowski et al. 2000; Zollner et al. 2000). These genes, however, are present within the accessory genome and as such are variable between the strain backgrounds of *S. aureus* (McCarthy and Lindsay 2013; Grumann et al. 2014).

Table 6-3 Chi2 comparison of immune-modulatory gene prevalence in cases and controls. * indicates p <0.05, ** indicates p value <0.01.

Immunomodulatory gene	Cases (%)	Controls (%)	Difference (%)	95% Confidence interval for difference	P- value
Hla (Haemolysin alpha)	100	100	0	N/A	N/A
Hld (Haemolysin delta)	100	100	0	N/A	N/A
*Psmb1 (*phenol soluble modulin beta 1)	100	100	0	N/A	N/A
Psmb2 (phenol soluble modulin beta 2)	90	71	19	3% to 34%	0.019*
SCIN (staphylococcal complement inhibitor)	96	98	2	-5% to 9%	0.57
sak (staphylokinase)	86	94	8	-4% to 20%	0.19
CHIPS (chemotaxis inhibitory protein)	54	40	14	10% to 45%	0.003**
entA (enterotoxin A)	36	26	13	-9% to 28%	0.31
entB (enterotoxin B)	14	4	10	-1% to 21%	0.086
entC (enterotoxin C)	12	31	19	3% to 34%	0.024*
entD (enterotoxin D)	6	6	0	-9% to 10%	0.98
entG (enterotoxin G)	42	73	31	13% to 50%	0.0015**
entH (enterotoxin H)	18	6	12	-1% to 24%	0.07
entJ (enterotoxin J)	6	6	0	-9% to 10%	0.98
entL (enterotoxin L)	42	76	34	15% to 52%	0.0007**
lukD (leukocidin D)	64	33	31	13% to 50%	0.0018**
lukE (leukocidin E)	64	33	31	13% to 50%	0.0018**
tst (toxic shock toxin)	10	26	16	2% to 31%	0.033*
eta (exfoliative toxin A)	6	6	0	-9% to 10%	0.98
ssl 1 (superantigen like protein 1)	44	71	27	9% to 46%	0.0058**
ssl 2 (superantigen like protein 2)	74	35	39	21% to 57%	0.0001**
ssl 3 (superantigen like protein 3)	70	29	41	24% to 59%	<0.00001**
ssl 4 (superantigen like protein 4)	62	29	33	15% to 52%	0.0008**
ssl 5 (superantigen like protein 5)	76	33	43	26% to 61%	<0.00001**
ssl 6 (superantigen like protein 6)	28	8	20	5% to 34%	0.0105*
ssl 7 (superantigen like protein 7)	72	37	35	17% to 54%	0.0004**
ssl 8 (superantigen like protein 8)	74	35	39	21% to 57%	0.0001**
ssl 9 (superantigen like protein 9)	76	41	35	17% to 53%	0.0004**
ssl 10 (superantigen like protein 10)	72	33	39	21% to 57%	0.0001**
ssl 11 (superantigen like protein 11)	20	6	14	1% to 27%	0.04*

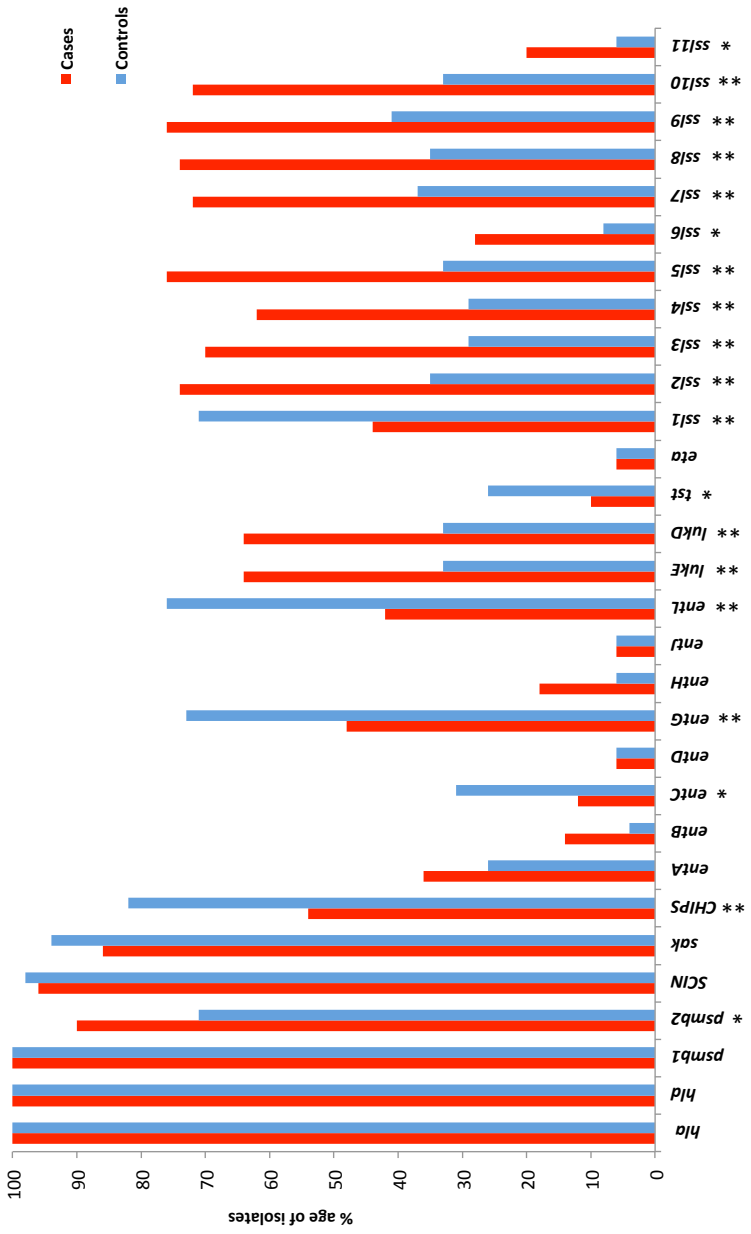

Figure 6-2 Immuno-modulatory gene prevalence in cases and controls. Comparison of gene presence between cases and controls of 30 *S. aureus* virulence factors as percentage of isolates. **psm**- phenol soluble modulin; **hl**- haemolysin; **SCIN**- staphylococcal complement inhibitor; **sak**- staphylokinase; **ent**- enterotoxins; **luk**- leuckocidins; **eta**- exfoliatin A; **tst**- toxic shock toxin; **ssl**- superantigen like proteins. * indicates p <0.05, ** indicates p value <0.01 (as per Table 6-3).

This genetically diverse isolate collection included more than 12 clonal backgrounds with evidence of partitioning into case and control groups. Therefore, the prevalence of these virulence genes was further assessed in the context of the clonal population structure of the isolate collection that differentiated disease groups. Presented in Figure 6-3, is a heat map of the presence or absence of these genes in comparison to the core phylogeny of all of the isolates. As shown by the phylogeny, isolates from respective clonal complexes fall into distinct clades. This demonstrates the underlying clonal influence on the previously observed distribution of the virulence determinants between cases and controls. The increased prevalence of TST in control isolates in comparison to cases (26% vs. 10%) (p=0.033; Table 6-3) can be explained by its carriage primarily by CC30 isolates, which was the most common clonal background in nasal controls. Similarly, enterotoxins G and L are more abundant in controls at 73% and 76% (p= 0.0015; Table 6-3) vs. cases at 48% and 42% (p= 0.0007; Table 6-3) respectively, because of their presence in CC30, CC45 and CC5 isolates which are more frequent in controls. Enterotoxins B and H are more frequent in cases (14% and 18%) (p=0.086; Table 6-3) compared with controls (4 and 6%) (p=0.07; Table 6-3) accounted for by their presence in CC1 and 59 isolates which are more frequent in cases.

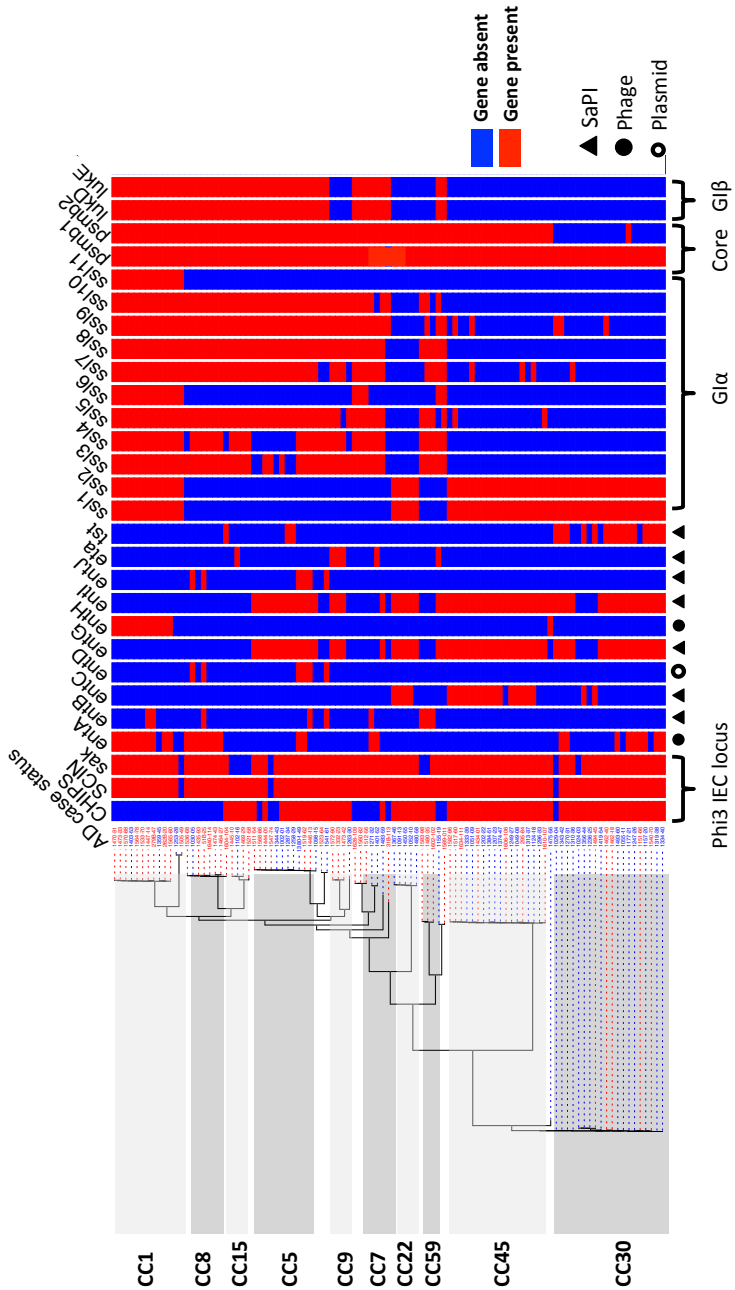

Figure 6-3 Immuno-modulatory gene prevalence in cases and controls according to clonal complex. The maximum likelihood core phylogeny is displayed along the left side. Shaded in alternating grey are the dominant clonal complexes (CC) between cases and controls. Taxon labels are coloured according to case status (AE-case red; Control-blue). Heat map demonstrates presence or absence of gene on basis of SRST2 method, with red indicating gene is present and blue absent. Gene names are indicated at top of heat map. The genomic regions or elements conferring virulence determinants are indicated by shapes or annotation of the heat map: SaPI- *S. aureus* pathogenicity island; GI- genomic island; IEC- immune evasion complex.

The superantigen-like proteins have varying reported functions including T cell stimulation and complement inhibition. One previous study reported *ssl8* and *11* to be more prevalent in AE disease isolates (Rojo et al. 2014). The data from this work suggests that all of the *ssl* genes with the exception of *ssl1* (Figure 6-2; Table 6-3) are more prevalent in case isolates. The *ssl* locus (within GIα) is known to be variable on a lineage dependent basis, both in the content of each *ssl* gene but also in their allelic variants. CC5 and 30 backgrounds, for instance, are known to lack carriage of an *ssl5* equivalent, whilst CC22, 30 and 45 lack *ssl8* (McCarthy and Lindsay 2013) The number of alleles of each gene is variable ranging from 1 in the instance of *ssl8* through to 13 in *ssl3*. Sequences of these genes from the same CC have been shown to have high levels of shared identity, whereas those from differing CCs have approximately 90% identity at amino acid level (McCarthy and Lindsay 2013). Variance between the disease groups can partly be attributed to clonality. The reference locus used for this comparison derived from MSSA476 belonging to CC1, therefore the variance is also indicative of sequence variance between homologues of these genes across other CCs, meaning they were not detected using SRST2. Therefore, it does not exclude the possibility of variants of these genes differing between disease and carriage populations. Finally, the leukocidins D and E (co-carried on GIβ) are almost twice as prevalent in cases (64%) compared with controls (33%) (p=0.0018; Table 6-3), but this element is notably absent in the CC30 and 45 strain backgrounds of dominating the control group (McCarthy and Lindsay 2013).

6.6.2 Antimicrobial resistance genes

Given the association between *S. aureus* and disease flares in atopic eczema, this patient group frequently receive treatment interventions in the form of antimicrobials and antiseptics. These isolates were therefore assessed to compare the prevalence of AMR determinants, which may subsequently impact upon response to therapy in AE cases.

All 99 isolates were assessed for the presence of antimicrobial resistance genes to common therapeutic agents, using SRST2. The mapping data was also assessed for each isolate to look for the presence of point mutations conferring resistance (Figure 6-4).

This comparison revealed β-lactamase carriage was almost identical between cases and controls, at 96% and 98% (p=0.57; Table 6-4) each reflecting their widespread distribution in the *S. aureus* population generally. Resistance determinants to fusidic acid (*fusA/B/C*) were identified in 38% of cases versus 14% of controls (p=0.0073; Table 6-4). Determinants of resistance to antiseptics (*qac* genes) were also more frequent in cases at 16% compared with 2% of controls (p=0.016; Table 6-4). Erythromycin resistance markers (*ermA/C/33*) were found in 14% of cases compared with 8% of controls (p=0.36; Table 6-4). Spectinomycin (*AAD9/ ant9*) resistance determinants were present in 14 % of cases versus 8% of controls (p=0.36; Table 6-4), whilst kanamycin resistance markers (*aphA-3/aacA/ aaca-aphD*) were all also more frequent in cases than controls (8% vs. 2%) (p=0.16; Table 6-4). Tetracycline resistance (*tetK/M*) determinates were found more frequently in controls (10%) compared with cases (4%) (p=0.23; Table 6-4). Methicillin resistance determinant *mecA* was found in single instances in both case and control groups, as was the point mutation in *ileS-1* associated with resistance to mupirocin (p=0.99; Table 6-4).

Table 6-4 Chi2 comparison of antimicrobial resistance determinant prevalence between cases and controls. * indicates p <0.05, ** indicates p value <0.01.

Antimicrobial resistance (genetic determinant)	**Cases (%)**	**Controls (%)**	**Difference (%)**	**95% Confidence interval for difference**	**P- value**
Penicillin (*blaZ*)	96	98	2	-7% to 5%	0.57
Fusidic acid (*fusA*/*B*/*C*)	38	14	24	7% to 40%	0.0073**
Antiseptics (*qacA*/*B*/*C*/*J*)	16	2	14	3.1% to 25%	0.016**
Erythromycin (*ermA*/*C*/ 33)	14	8	6	-6% to 18%	0.36
Spectinomycin (*AAD9*/ *ant9*)	14	8	6	-6% to 18%	0.36
Kanamycin (*aphA-3*/ *aacA*/ *aaca-aphD*)	8	2	6	-2% to 15%	0.16
Tetracycline (*tetK*/*M*)	4	10	6	-4% to 16%	0.23
Methicillin (*mecA*)	2	2	0	-6% to 6%	0.99
Mupirocin (*ileS-1* V588F)	2	2	0	-6% to 6%	0.99

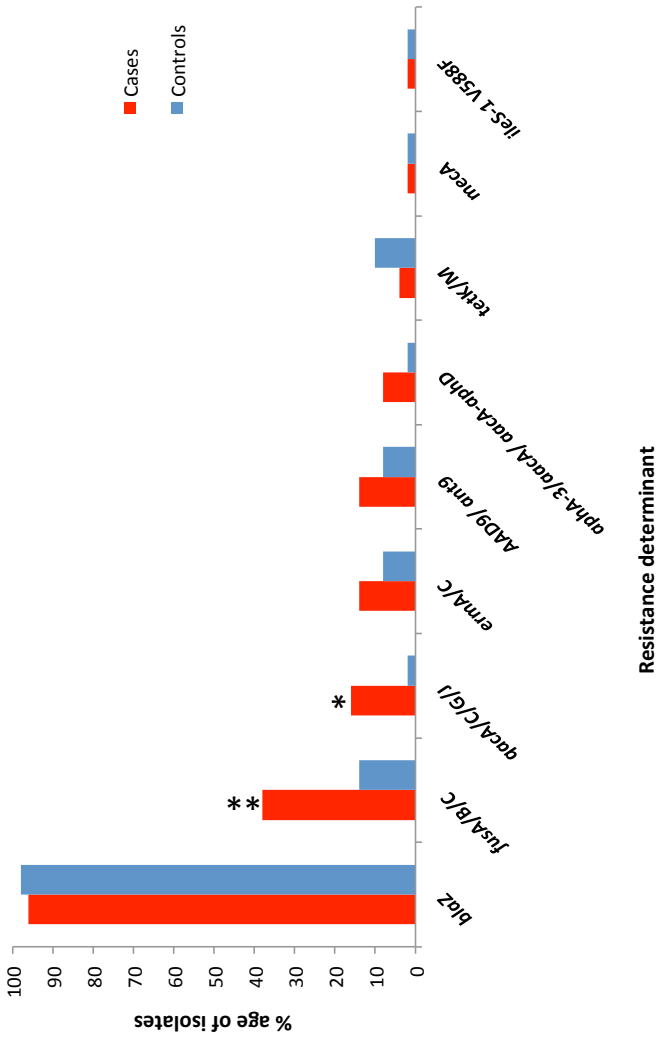

Figure 6-4 Prevalence of antimicrobial resistance determinants across case and control populations. * indicates p <0.05, ** indicates p value <0.01 (as per Table 6-4).

Topical fusidic acid therapy is probably the most common antimicrobial intervention in AE, predominantly in the community setting. Both plasmid- derived and point mutation associated resistance determinants were found in cases and controls. Within the case isolates, 4% were found to be carrying *fusB*, in comparison to 2% of controls (p=0.57; Table 6-5), whilst *fusC* was identified in 20% of case isolates compared with 10% of controls (p=0.17; Table 6-5). Acquired mutations in *fusA* were significantly more common in cases, being found in 14% of cases compared with 2% of controls (p=0.029; Table 6-5). Four mutations were identified including: *fusA* L461S (6% cases; 0 controls), *fusA* L461K (4% cases; 2 controls), *fusA* A376V (2% cases; 0 controls) and *fusA* H457Y (2% cases; 0 controls). These *fusA* point mutations have been reported to confer high-level fusidic acid resistance (MIC >32µg/ml), in comparison (Chen et al. 2010)**,** in comparison to carriage of *fusB* or *fusC,* which are associated with lower levels of resistance (MICs <32µg/ml) have been shown (O'Neill et al. 2007; Chen et al. 2010).

Table 6-5 Chi2 comparison of Fusidic acid resistance determinant prevalence between cases and controls.

Fusidic acid resistance (genetic determinant)	Cases (%)	Controls (%)	Difference (%)	95% Confidence interval for difference	P-value
fusA	14	2	12	2% to 22%	0.029
fusB	4	2	2	-5% to 7%	0.57
fusC	20	10	10	-4% to 20%	0.17

Carriage of the small plasmid-associated *qac* genes, which are associated with reduced susceptibility to antiseptics such as chlorhexidine, is of specific relevance in cases because of frequent use in AE. A total of 16% of cases were carrying a *qac* gene, *qacA* (4%) (p=0.16; Table 6-6), *qacC* (10%) (p=0.023; Table 6-6) or *qacG* (2%) (p=0.31; Table 6-6), whilst a single control was carrying *qacJ* (2%) (p=0.31; Table 6-6). Previous studies have reported varying rates of carriage of these genes, with *qacA* found in 5.7% of isolates, *qacC* at 0.3% and *qacG* at 0.1% (Furi et al. 2013) in isolates from the UK of community and hospital origins.

Table 6-6 Chi2 comparison of antiseptic resistance determinant prevalence between cases and controls.

Antiseptic resistance (genetic determinant)	Cases (%)	Controls (%)	Difference (%)	95% Confidence interval for difference	P-value
qacA	4	0	4	-1% to 9%	0.16
qacC	10	0	10	2% to 8%	0.023
qacG	0	2	2	-2% to 6%	0.31
qacJ	0	2	2	-2% to 6%	0.31

Tetracycline resistance determinants were found in both cases and controls at rates of 4% and 10% respectively (p=0.23; Table 6-4). This is interesting as all of the participants in this study would were below the minimum age recommended for safe prescribing of tetracyclines of 12 years (BNF for Children 2016) and therefore is likely representative of prescribing practices in the wider population.

6.6.3 Adhesion genes

Individuals with AE are exceptionally prone to colonisation by *S. aureus*. Rates of colonisation of 70-90% have been reported for more than 30 years in clinical dermatology (Leyden et al. 1974; Hauser et al. 1985; Hoeger et al. 1992). These high rates of colonisation have been hypothesised to occur as a consequence of multiple aetiological factors involved in AE. Experimental evidence has however also demonstrated increased binding of *S. aureus* specifically to AE-affected skin in comparison to those affected by the common inflammatory disease psoriasis, and normal individuals (Cho et al. 2001a).

The surface proteins of *S. aureus* are directly mediating contact with the host surface and colonisation making them an important class of virulence determinants relevant to this disease state. All 99 isolates were assessed for their content of cell surface genes in order to look for differences in their presence between disease states. These genes fall across both core and

core variable genomic regions meaning they range from highly conserved to variably present across CCs (McCarthy and Lindsay 2010).

Bioinformatic characterization of the surface proteins from short read paired-end sequencing is challenging. These genes have repetitive structures, meaning their sequences are of low complexity and can be difficult to obtain sequence reads across. They are also specifically prone to the accumulation of mutations and undergo recombination. This means the methods chosen to identify their presence or absence can be inaccurate because of lack of homology to a reference allele. Finally, these genes frequently have sequences of more than 2kb, which, in combination with repetitive sequences, makes obtaining the sequence for the entire gene challenging with the short length reads from MiSeq sequencing. The following results are therefore representative of a combination of bioinformatic methods and approaches that provide a population overview of the isolates between cases and controls within the limits of the data.

A reference allele database composed of sequences of *S. aureus* surface proteins was complied (Materials and Methods- Chapter 2; Table 2-14) and isolates were mapped to this using BWA (Li and Durbin 2009) initially. The initial results obtained using this method demonstrated that it failed to detect the presence of genes that have previously been reported as being highly conserved. Figure 6-5 is an example of this issue when trying to detect Extracellular adhesion protein (*eap*), an MHC class II analogue protein thought to play a role in *S. aureus* internalisation into keratinocytes (Bur et al. 2013). Using this method suggested that the gene was entirely absent from CC30 and 45 isolates, despite being shown as present in reference genomes of 25 CCs previously (McCarthy and Lindsay 2013). Clonal complex 8 isolates mapped across the entire reference allele, whilst there was variable coverage from isolates of the other clonal complexes. A subsequent BLAST (Altschul et al. 1990) comparison of alleles from CC1, 30, and 45 reference genomes to the CC8 allele revealed that they shared

84% sequence identity at DNA level. This indicated that for further comparison a method incorporating multiple reference alleles per gene representative of the major clonal backgrounds including CC5, CC8, CC22 and CC30 were required. As an alternative approach each of the isolates was then mapped against a reference allele, composed of clonal complex specific variants of the gene using SRST2 (Inouye et al. 2014). A limited number of these genes have been studied in AE therefore the adhesin genes were chosen on the basis of known function in colonisation (Foster et al. 2014).

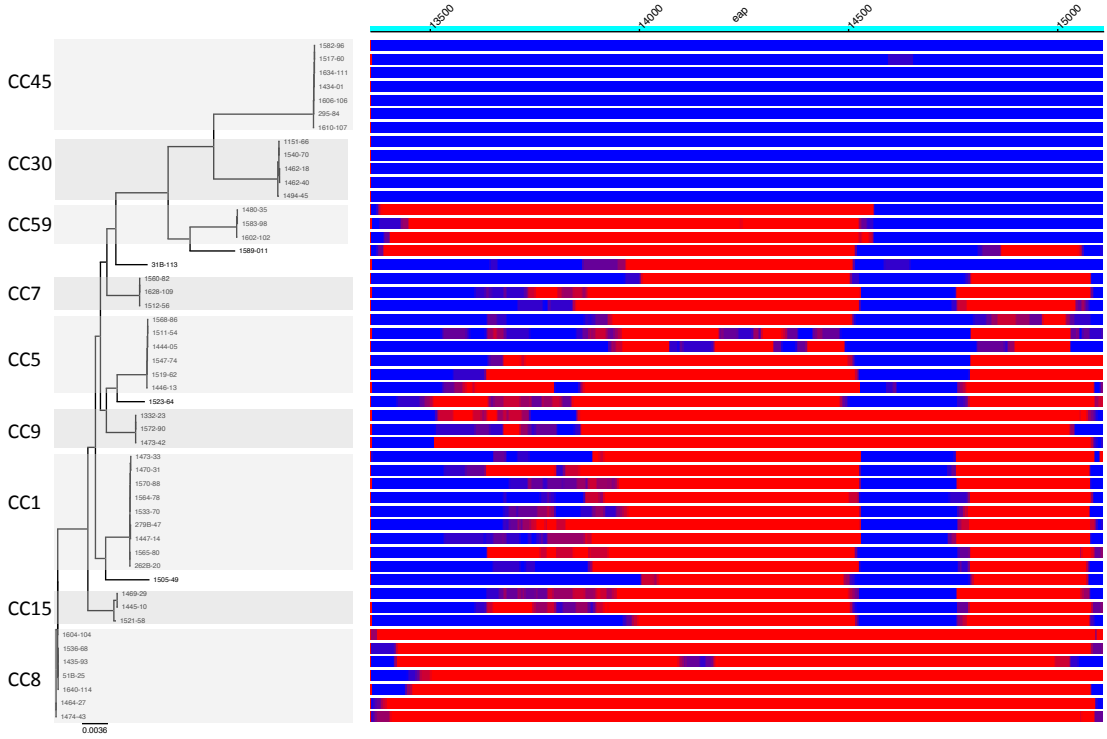

Figure 6-5 Mapping against reference allele of *eap* (extracellular adherence protein) from CC8 reference genome NCTC8325 (locus ID: SA0UHSC_02161) in Irish AE cases. Heat map generated using iCANDY illustrating mapped reads across reference allele in 50 AE cases. The relative chromosomal position in accordance to *S. aureus* NCTC8325 (accession: CP000253) is indicated at top of figure. Core phylogeny is presented at the left of the figure, with shading depicting respective clonal complexes (CC) of isolates. Red indicates the presence of mapped reads across the gene at a depth of coverage greater than 10. Blue indicates absence of mapped reads to the reference.

The results of this analysis are demonstrated in Table 6-2, and represent the gene content of this class across both disease groups. There were no examples of variable presence of these genes between cases and controls of the same clonal background. Instead, what is again

evident is the clonal influence on the gene content of isolates in both cases and controls. Identified in all of the isolates were *spa* (immunoglobulin binding protein A), *isdA*/ *B* /*H* (iron regulated surface proteins), *eap, clfA* /*B (*clumping factors) and *fnbpA* (fibronectin binding protein).

Within the *S. aureus* genome there are loci where surface proteins are found together, and along with their repetitive nature they are prone to recombination (Basic-Hammer et al. 2010). For this reason there is variability between CCs in these protein both in their sequence as well as their presence. For instance *fnbpB* (fibronectin binding protein B) which mediates adhesion to host extracellular matrix (Foster et al. 2014), was not detected in the CC22 or 30 isolates in cases or controls. This fits with previous characterisation of the ST22 (EMRSA15) and ST36 MRSA (MRSA252) lineages which are known to have a fusion of *fnbpA* and *B* arising from recombination event, as opposed to two separate copies of the fibronectin binding genes (Holden et al. 2013). The serine aspartate repeat proteins, *SdrC* and *D*, are known to bind to desquamated epithelial keratinocytes (Corrigan et al. 2009) and were also variably present on a lineage dependent basis. Both were absent in CC121, whilst *SdrD* was also not found in CC30 and 59 by this method.

Table 6-7 *S. aureus* surface protein profile of Irish AE cases and controls. Shading indicates presence of these genes in the clonal backgrounds of both the case and control isolates. (Alternate grey and blue shading is used to distinguish columns relating to clonal complexes; blank cells indicate gene was not detected).

	Clonal complex of isolates													
Gene	**1**	**5**	**7**	**8**	**9**	**15**	**22**	**25**	**30**	**45**	**59**	**121**	**779**	**1290**
spa														
IsdA														
IsdB														
IsdH														
eap														
clfA														
clfB														
fnbpA														
fnbpB														
sdrC														
sdrD														
sasC														
sasD														
sasG														
cna														

Two examples of surface proteins that differed in their presence in the dominant CCs (and STs) differentiating the disease groups were *SasG* (*S. aureus* surface protein G) and Cna (collagen adhesin). Surface protein G was found in CC1, CC5, CC8, CC15, CC59, ST779 and ST1290 isolates which represent the CCs present more frequently in the cases with the exception of CC5. The collagen adhesin gene, which has was found in isolates of CC1, 22, 121, as well as the dominant control strain backgrounds CC30 and 45. Both genes have been shown to be involved in colonisation, with *sasG* directly adhering to desquamated epithelial cells (Corrigan et al. 2007), whereas *cna* has been demonstrated to confer enhanced ability to colonise the corneal epithelium (Rhem et al. 2000). Whilst *sasG* has been previously reported as being significantly increased in AE disease isolates previously (Rojo et al. 2014), the results of this study suggest the clonal background of the AE case isolates accounts for this finding.

6.7 Pangenome comparison of disease and carriage isolates

The previous analyses used a candidate gene approach and demonstrated differences between cases and controls that may be explained by the underlying genetic makeup of the isolate collection. In addition to surface proteins there are likely to be other sets of lineage-specific genes that may be relevant in their association with disease. Using WGS allows a broader assessment of the gene content beyond candidate genes, extending the analyses to compare the pangenome content of disease and carriage isolates looking for signals of association with disease. This approach has been previously used successfully to identify genetic factors that distinguished carriage and invasive isolates in the pig pathogen *Streptococcus suis* (Weinert et al. 2015)

The whole genome data set for all 99 isolates was compared to look for evidence of genome contents associated with disease. To do this the annotated genome assemblies of the isolates were used to compile a single pangenome using Roary (Page et al. 2015). In brief, this method extracts protein-coding regions (CDSs) from the input assemblies, converting them to protein sequences, which are then filtered to remove incomplete sequences. A BLASTP comparison is made against all the sequences, and sequences are finally clustered into homology groups on the basis of 95% sequence identity or more. This allows clustering of genes in order of their presence from highly conserved (core) to variably present (accessory). The output is a database of homology groups (genes) identified amongst the isolate representing core and accessory genome components. Homology groups (HGs) conserved amongst all isolates were removed, leaving HGs representative of core variable and accessory genome components. These were converted to a binary matrix on the basis of presence or absence and used as the input for Discriminant Analysis of the Principal Components (DAPC) (Jombart et al. 2010). This modified form of principal component analysis was then used to look for evidence of gene content with association to disease status.

The preliminary results from this analysis suggested a clear differentiation between groups. From the DAPC analysis the genes identified as differentiating the groups were then assessed. This revealed examples of core genes such, as *isdH*, which were predicted to be differentially present in case and control populations. Given the previous work done on the isolate collection indicates that this result was incorrect, as mapping methods had demonstrated this gene to be conserved across isolates. Further assessment revealed that this gene was identified twice as separating the populations on disease status. This suggests two possibilities, firstly that there was an allelic variant differentiating the isolates across the study populations, or secondly that there was fragmented sequence for this cell wall anchored protein and during assembly of the pangenome, the fragmented sequences had been clustered into separate homology groups because of BLASTP similarities of less than 95%. Proteins such as isdH have centrally repetitive sequence regions meaning that they are frequently not assembled in to a single coding sequences (CDS). Comparison of the annotated assemblies demonstrated truncated CDSs for this gene, meaning the result was an artefact of the assembly quality.

Importantly this highlighted the need to further assess the quality of the input assemblies when using this method. A comparison of the genome assembly quality was then made for all of the isolates which showed that over 56 of the isolates had more than 100 contigs, and of these 6 had more than 400 contigs (Figure 6-6 A). The fragmented quality of some assemblies would therefore mean considerably higher numbers of broken protein coding sequences, which in turn would influence the distribution of clustered proteins and therefore their relative presence or absence. This was illustrated in the number of genes present within the pangenome assembly (Figure 6-6 B), which was much greater than would be expected at more than 6381.

In order to improve the quality of the data set for pangenome analysis, assemblies with contigs of greater than 400 were removed. The remaining 93 assemblies were then used to produce a new pangenome assembly with Roary. This led to a reduction in the total number of genes by 215 from 6381 to 6166. Of this total number of genes, 1542 were conserved amongst all isolates and were removed. The remaining 4624 genes, representing core variable and accessory genome components, were used to compile the binary matrix of presence/ absence for use in the repeat DAPC analysis.

Figure 6-6 Assessment of genome assemblies and number of genes in pangenome assessment. A- graph illustrating number of contigs across 99 genome assemblies. Red line indicates contigs over 400. **B**- Total number of genes in pangenome. Solid black line indicates number of conserved genes across all 99 *S. aureus* genomes used. Perforated line indicates total number of genes (inclusive of core, core variable and accessory genes).

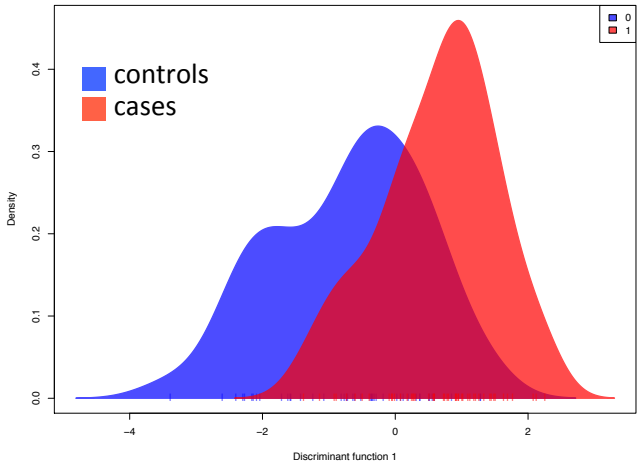

Figure 6-7 Discriminant analysis of Principal components applied to 93 *S. aureus* isolates (49 cases and 44 controls using presence/ absence data for genes in the accessory genome.

The DAPC using 93 isolates revealed considerable overlap in the accessory genome components shared between the isolates of the two study groups (Figure 6-7). This indicates the lack of genetic difference between isolates from AE cases and controls. The genes accounting for the limited differences between cases and control populations are demonstrated in Figure 6-8 and presented in Table 6-3. This provided examples of potentially relevant genes differentiating the disease groups. For instance, the transposase found in 22 cases compared with 11 controls (explanatory variable 1037;Figure 6-8, Table 6-3) is an example of differential carriage of a possible virulence determinant. Of the 22 isolates in the case population carrying this gene, 16 were in clonal backgrounds more frequent in cases including CC7, CC8, CC9, CC15 and CC59. A similar finding is noted in the example of enterotoxin O (*seo*) (explanatory variable 709; Figure 6-8, Table 6-3) found in 34 controls compared to 19 cases. Of the control samples carrying *seo,* 31 are from the clonal complexes commoner in controls including CC5, 22, 25, 30 and 45. These results therefore again represent the clonal influence of this small but diverse collection of isolates rather than candidate genes associated with disease status.

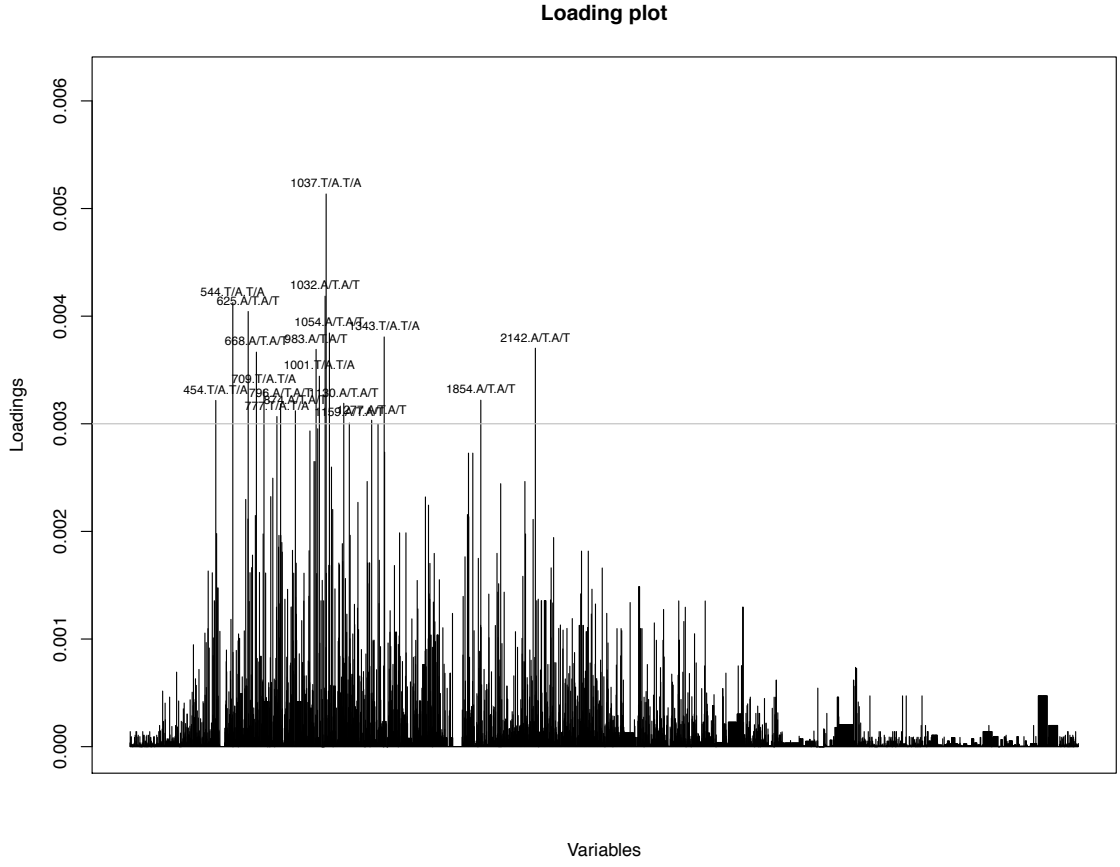

Figure 6-8 Accessory genes differentiating case and control population. Peaks indicate homology groups (genes) separating case and control populations using DAPC (Descriminant analysis of principal components). Numbers indicate explanatory variable, which corresponds to a homology group within the pangenome produced from the 93 isolates.

The cell wall anchored protein *sdrD* (explanatory variable 1343; Figure 6-8, Table 6-3) was another gene predicted to differ in presence between the two disease states. With these methods, 21 isolates were predicted to have the gene, 14 cases and 7 controls. This was not in congruence with results from the cell surface protein analysis in Table 6-2, which demonstrated that whilst variably present in clonal complexes, it was widely distributed in the isolate collection. This result was found to be due to a fragmented gene, however, only representing a C-terminal fragment, with another three *sdrD* homology groups being present within the pangenome, thus demonstrating the assembly quality issue.

Table 6-8 Genes differentiating cases and controls from DAPC analysis. Explanatory variable number corresponds to peak in plot in Figure 6-8. Explanatory variables corresponding to homology groups within the pangenome of the 93 case and control isolates.

Explanatory variable	Total number of isolates	Number of cases	Number of controls	Homology group/ Gene annotation in Pangenome
454	80	39	41	*vraD* ; ATP binding / ATPase activity
544	72	31	41	*ylbE;* oxidoreductase
625	61	28	33	fic/DOC family protein
668	57	24	33	*melR_2*; Arac family transcriptional regulator
709	53	19	34	*seo*; Enterotoxin O
796	48	17	31	Putative membrane spanning protein (DUF443 superfamily)
874	41	29	12	DUF576 family protein
983	35	18	17	Hypothetical protein
1001	35	14	21	Phage protein
1032	33	15	18	DUF600 family protein
1037	33	22	11	Transposase
1054	32	18	14	Putative PVL-like protein
1130	28	11	17	Hypothetical protein
1159	27	12	15	Putative ATP binding protein
1277	22	17	5	*ylbE;* oxidoreductase
1343	21	14	7	*sdrD*; Serine-aspartate repeat containing protein D
1854	14	7	7	Phi ETA orf 22-like protein
2142	10	5	5	Phage protein

The oxidoreductase *ylbE* (explanatory variables 544 and 1277; Table 6-3 and Figure 6-8) was present in all isolates, but differentiated into two separate homology groups. Again this suggested the possibility of an allelic variant differentiating the disease states. However BLAST comparison of the sequence of these two homology groups revealed that they shared 99% sequence identity. Given the thresholds for clustering into homology groups with Roary, these genes should therefore not have been separated. This again suggests a methodological artefact has occurred.

The results presented in the earlier sections examining toxins, AMR genes and adhesin genes did identify differences between cases and controls. Albeit these are influenced by the clonal background of the isolates, their absence in the output from this analysis again indicated that the results from this pangenome study were not reliable.

6.8 Discussion

This analysis has demonstrated clear differences between the population structure of isolates from cases and controls. Using a WGS approach allowed wide assessment of the genetic content between disease and carriage isolates which has not been done in previous studies undertaking similar comparative analysis. Differing gene content between the disease populations reflects the clonal backgrounds of the samples, but nonetheless indicates that there are differences of biological relevance in the context of disease-associated isolates.

Population structure of colonising strains

Numerous studies have compared AE disease and carriage and carriage isolates to associate genetic content with disease causation in AE (Capoluongo et al. 2001; Kim et al. 2009; Yeung et al. 2011). These studies have concluded that there is no prevailing *S. aureus* clone associated with the disease. Our findings demonstrate several clonal backgrounds that are more prevalent in cases including clonal complex 1 (Figure 6-1). This strain preponderance was also observed in the AE cases recruited through the Dundee study presented in Chapter 5. Comparison of the strains in our cases with those isolated in a cohort of Korean AE patients revealed similarities (Kim et al. 2009). In this study 33.3% of isolates from AE cases belonged to CC1, which were found in 20% of the Irish cases (and 44.4% of Dundee cases in Chapter 5). The comparatively lower rates of CC30 isolates in AE cases compared with controls in this study (10.0% vs. 33.0%) (p=0.0058; Table 6-2) has also been reported previously (Kim et al. 2009; Yeung et al. 2011; Rojo et al. 2014). If nasal carriage studies are taken into consideration, they are by definition investigating colonisation of a niche site, and repeatedly show the CC30 and 45 lineages are dominant (Melles et al. 2004; Monecke et al. 2009). This is reflected in the nasal carriage isolates in this study and those in Chapter 4. Comparatively, clonal complex 1 isolates are generally noted with much lower prevalence in nasal carriage (Melles et al. 2004;

Monecke et al. 2009). These findings raise the question of whether there are lineage specific features that make strain backgrounds more adept at colonisation of the very different cutaneous environments of the nasal epithelium compared to inflamed skin. This highlights an area where future studies could be directed.

Toxin gene content

A defining feature of AE is T-cell mediated cutaneous inflammation, and as such much of the previous research in this disease pertaining to *S. aureus* has been centred on toxins, which stimulate T cell responses. The presence and production of toxins have frequently been assessed in AE disease isolates, with the enterotoxins (A to D) and TSST being the most widely studied examples (Bunikowski et al. 2000; Zollner et al. 2000; Schlievert et al. 2010; Pascolini et al. 2011; Rojo et al. 2014). Prior studies reporting on toxin gene prevalence and association to AE disease severity have tended not to correlate the results with strain backgrounds of isolates (Pascolini et al. 2011; Rojo et al. 2014). This makes the results of such studies difficult to interpret on the basis of the known clonal variability of the studied toxins genes (Holtfreter et al. 2007). In this study, if we were simply to compare the prevalence of TST gene presence across the disease populations it would appear to be negatively correlated with AE disease, but this result can be explained by the genes presence in 26% of controls versus 10% of cases ($p=0.033$; Table 6-3) on the basis of its preferential carriage by CC30 isolates (Figure 6-2) (Holtfreter et al. 2007). Similarly enterotoxins A and B in this context are more frequent in cases (Table 6-3) but these findings are CC related, with both toxins being more frequent in clonal backgrounds over-represented in cases including CC1, 8, and 59 (Figure 6-2). In a more recent study of AE disease-associated isolates from Spain, the authors suggested the possible clonal influence relating to their detection of toxins being more frequent in cases than controls (Rojo et al. 2014). The findings from this work demonstrate that this is the case and provide an alternate explanation for the findings of previous work with respect to gene prevalence. The results presented in Table 6-3 demonstrate why comparative gene approaches are potentially

misleading in the context of disease association in AE. Where differences in the prevalence of certain genes were shown to be very statistically significant in this study, for instance the leukocidins D and E, as in the study by Rojo *et al.* (2014), this analysis signals to a more important consideration, that is the overall difference in strain backgrounds.

The candidate gene approach has yielded strong functional evidence linking *S. aureus* to the pathogenesis of atopic eczema, the phenol soluble modulin, delta haemolysin, is a good example of this. It was recently shown to directly trigger atopic type skin inflammation through mast cell degranulation in murine skin (Nakamura et al. 2013). As shown in Figure 6- 2, this gene was detected in all case and control isolates, demonstrating its conservation as part of the core genome. Another toxin of this class is alpha haemolysin, which was also conserved in all of the isolates in this study. The observed conservation of these genes highlights an important limitation of the method adopted in this work, which is the inability to determine if they are actually expressed or not. Highly conserved genes may conversely be significant and relatively unexplored avenues in AE research. For instance there is evidence of lineage specific variation in secretion of haemolysin alpha in vitro (Monecke et al. 2014b). The authors of this study demonstrated low level or no toxin production in isolates belonging to CC2, CC30, CC45 or CC59. While isolates from CC1, CC5, CC8 and CC15 produced variable to high levels of the toxin (Monecke et al. 2014b). These results are interesting as they bear some correlation to the disease group in this study, where higher levels of production of the toxin were found in lineages more frequent in AE cases including CC1 and CC8. These results by no means address the complex nature of regulation of expression of toxins such as haemolysin alpha, which are frequently controlled by multiple regulatory systems (Montgomery et al. 2010). They do raise the interesting question of whether other toxins vary in their excretion relating to their clonal background, which would be an important issue to address in AE research given that one of

the main findings of the works in Chapters 4,5 and 6 are strain preponderances between carriage states.

Adhesin genes

The inherent ability of this organism to adhere to host epithelial surfaces is clearly very relevant in eczema, where reported colonisation rates range between 70 and 90% (Leyden et al. 1974; Warner et al. 2009). Few studies have extended characterisation of AE associated isolates with respect to surface proteins mediating adhesion, beyond detection of a few candidates including *fnbpA*/*B*, *cna*, *clfA* (Lomholt et al. 2005; Pascolini et al. 2011). The results in Table 6-2 demonstrate that whilst many of these genes are conserved across lineages, such as *fnbpA*, *clfA*/*B*, there is also lineage specific variation. Surface protein G is an example of this, being absent in lineages more frequently identified in carriage controls, CC30 and CC45. The product of this gene is involved in adherence to keratinocytes (Foster et al. 2014) and was recently reported as being associated with AE disease severity (Rojo et al. 2014). The case isolates reported by that group were predominantly from CC5 and CC15, which as demonstrated by this work are backgrounds with *sasG* as a component of their core genome. The lower prevalence of *sasG* in controls in this study can also be explained by the predominance of CC30 and CC45 isolates (Rojo et al. 2014), just as is noted from this work. The collagen adhesin, reported to confer enhanced survival in blood as well as colonisation (Foster et al. 2014) in comparison seems to negatively correlate with AE disease isolates, being absent in backgrounds found more frequently in AE including CC7, CC8, CC9, CC15 and CC59. These indicate a potentially relevant genetic difference in AE associated isolates from this study, again emanating from the apparent strain segregation between cases and controls.

There has been relatively limited study of *S. aureus* surface proteins and their relevance in AE. Their repetitive nature and size make characterisation and even detection from sequence data difficult. In order to simply detect presence of these genes in this work required comparison to

multiple alleles per gene given their clonal variance in both sequence and presence (McCarthy and Lindsay 2010; Foster et al. 2014). It has been suggested that variation in certain surface proteins such as the fibronectin binding proteins changes the ligand binding affinity of these proteins (Foster et al. 2014). During this analysis detection of genes encoding the clumping factors, despite being highly conserved was difficult because of sequence divergence between clonal backgrounds. The initial approach of comparison to a single reference allele using SRST2 led to results suggesting the *clfB* gene was variably present in the 99 isolates. Subsequent comparison of the variation within the isolates to a single allele demonstrated that there was clonal sequence variation within regions of the gene. For instance comparison of all isolates against *clfB* allele from CC8 reference genome NCTC8325 (locus ID: SAHOUSC_02963) demonstrated sequence divergence in regions bp 591 to 1125, and from bp 1300 onwards in all isolates except those from CC8 background. These regions fall within reported binding domains N2/3 of the *clfB* protein, which are responsible for interaction with the human epidermal keratins 10 and loricrin (Walsh et al. 2004; Mulcahy et al. 2012). Whether such variability would impact upon binding affinities needs to be experimentally interrogated, but CC-related variance in adhesive properties of a conserved gene would be highly relevant in eczema. This demonstrates a further area of study required to advance our understanding of this organism's enhanced colonisation of atopic skin.

Antimicrobial resistance in AE and nasal controls

The strong clinical association between *S. aureus* and AE disease flares mean that antimicrobials are a frequent intervention in this patient group. This is evident from the resistance profiles between the two populations. Whilst carriage of a beta-lactamase was almost equal between cases and controls (96% vs. 98%; p=0.57, Table 6-4), fusidic acid resistance determinants were more than twice as frequent in cases compared with controls (38% vs. 14%; p=0.0073, Table 6-4). Rates of fusidic acid resistance in *S. aureus* isolates from dermatology patients in general in the UK are noted to be much higher than other patient

groups. One study has demonstrated rates of between 40% and 50% fusidic acid resistance in dermatology patients in comparison to other community or hospital based patients where rates were found to be 20% (Mitra et al. 2009). This reflects the frequent prescribing of this therapy for minor skin infections. The fusidic acid resistance rates in AE in this study are higher than reported in other studies where rates of between 5% and 30% are reported, but prescribing practices in the country of study will clearly influence these (Salah and Faergemann 2015; Park et al. 2016).

The plasmid conferred fusidic acid resistance determinants *fusB* and *C* were present in cases and controls, but were twice as common in cases. Acquired point mutations in *fusA*, were identified in 14% of cases compared with 2% of controls (p=0.0073; Table 6-5). The prevalence of plasmid-derived determinants could be accounted for by strain backgrounds in cases and controls, for instance in cases 70% of *fusC* genes are in CC1 (ST1) isolates. The *fusA* mutations are more suggestive of exposure to therapy and adaptive mutation. However this cannot be definitively concluded upon for these patients as this point mutation can arise spontaneously in *S. aureus* populations with a reported frequency of 1 in 10^6 to 10^8 CFU (Howden and Grayson 2006). These observations raise two interesting points for consideration. Firstly, has the antimicrobial resistance profile of the colonising strain potentially influenced the strain prevalence seen in cases in this collection, with fusidic acid resistance in CC1 and CC8 isolates being the primary example? Secondly, are the acquired point mutations in *fusA*, found almost exclusively in cases, the result of use of fusidic acid? Whilst a previous study has suggested that short term usage of fusidic acid in AE does not contribute to resistance, this work assessed the emergence of resistance in individuals using such preparations for 2 weeks (Ravenscroft et al. 2003). In clinical practice much shorter courses of this therapy such as 5 days are frequently prescribed. It is also not uncommon for patients to report returning to a therapy previously prescribed, as a rescue during flares, given previous benefit. Whether short bursts of this therapy such as this result in *fusA* mutations is an important clinical question in AE given that

they confer high level resistance and have the potential to drive SCVs which may persist intracellularly contributing to treatment failure (O'Neill et al. 2004; Sendi and Proctor 2009).

Topical antiseptics are another very common therapy in eczema, and primarily contained with emollients used as soap substitutes. Whilst they are generally used for clinically infected eczema, they are prescribed widely as a preventative therapy despite a lack of clinical evidence to support their use in this context (Bath-Hextall et al. 2010; Eichenfield et al. 2014). In AE cases 16% of isolates were found to be carrying one of the *qac* genes, *qacA*, *qacC* or *qacG*, whilst a single control carried *qacJ* (p=0.016; Table 6-6). Carriage of these plasmid encoded efflux pumps has been shown to confer increased minimum inhibitory concentrations (MIC) to compounds including benzalkonium chloride and chlorhexidine (Furi et al. 2013). These both constitute active compounds in regularly prescribed in emollients and soap substitutes in dermatology, such as Dermol 500®. The single control isolate carrying *qacJ* is of interest as this has been reported in association with veterinary isolates or equine origin (Bjorland et al. 2003), so this finding may reflect environmental exposure. A recent study investigated the prevalence of *qac* genes in contact lens wearers over a 6-month period whilst using low concentration antiseptic solution (Guang-Sen et al. 2016). They reported that whilst conjunctival carriage of *S. aureus* reduced there was increasing prevalence of carriage of *qacA* and *qacB* from 4.4% to 15.4% over the 6-month period, with the study authors suggesting that use of low dose antiseptics was selecting for carriage of strains with these determinants. Whilst MICs to antiseptics may only change modestly with carriage of these genes (Furi et al. 2013), this is still relevant to clinical practice in dermatology. Not only are the topical therapies containing these compounds (at low concentrations) likely to be less effective, they are also potentially as irritant to skin as the soap products they are substituting (Basketter et al. 2004).

Limitations of pangenome approach in this study

Assessment of the genetic content of these isolates for specific genes, and gene classes yielded potentially relevant differences in the context of AE disease. The DAPC analysis attempted to look for genetic content at the genome level however several attempts using this method didn't yield sufficiently reliable results to take further. From the outset of this study, the small number of isolates was always going to be a limiting factor. This genome level approach has proven to be very effective in other bacterial pathogens, but with data sets of almost four times this size of this one (Weinert et al. 2015). The quality of the genome assemblies used for the analysis had an impact upon the results obtained. Reduction in the number of assemblies used from 99 to 93 in subsequent repeats of the analysis removed the apparent observed association. This may reflect the a true lack of association on the basis of disease status in this small collection, but the assessment of the gene content by the other means described in this chapter did highlight differences in disease groups. Exactly why differential presence of genes such as examples of toxins or resistance genes were not discriminated by this method is still unclear, but there are likely to be a combination of factors at play. In some instances, this may be the result of the clustering into homology groups during pangenome construction, whereby genes with 95% or more sequence homology would be grouped together. Broken protein coding sequences may potentially have resulted in conserved regions of genes in the same class being clustered together.

6.9 Concluding remarks

The results of this study have demonstrated potentially relevant differences in disease and carriage associated *S. aureus* isolates. Evident from comparison of the populations is the apparent strain segregation on the basis of disease status, which reflects the results found in Chapters 4 and 5. Whilst candidate gene searching has yielded some interesting differences,

many examples in this work and previous studies can be explained by the clonal structure of the isolate collection, and is likely to be a confounding issue encountered in future studies. It importantly highlights that need for a broader approach to understand this organism in the context of eczema, where undertaking genome wide association studies on larger data sets would be an obvious progression. The impact of antimicrobial prescribing in this patient population is again evident from this sample collection. These results indicate where future clinical studies should be directed, not only to reduce the occurrence of antimicrobial resistance but also potentially the use of topical therapies outwith current recommended guidelines.

7 Conclusions and outlook

The work conducted in this study has provided new insights on the evolution of *S. aureus* from two very differing perspectives. First, from a population level approach, over a time scale of decades to trace the first ever MRSA. Second, over much shorter time scales, characterising within host diversity during colonisation and disease flares of AE in children. This work finally assessed the population structure of carriage versus AE disease associated *S. aureus* isolates, looking for evidence of genetic determinants associated with disease.

7.1 The selective pressures of clinical practice and the emergence of MRSA

Epidemiological evidence has always suggested that methicillin resistance in *S. aureus* arose one year after the introduction of methicillin into clinical practice in 1959, in response to the introduction of this novel antibiotic into clinical practice. By using WGS of a collection of 209 of the earliest MRSA isolates identified in Europe from between 1960 and the late 1989 it was possible to trace the evolutionary history of the first MRSA.

The results presented in chapter 3 demonstrated that *S. aureus* acquired an SCC*mec* element carrying the methicillin resistance determinant long before the drug was first used, and at time of increasing penicillin usage during the 1940s. When methicillin was introduced, it did not select for the *de novo* evolution of MRSA, rather it selected for the emergence of MRSA from a cryptic population that had already acquired the SCC*mec* element. Given that *blaZ*, which encoded a β-lacatamase that provides resistance to penicillin, was increasing in prevalence in the *S. aureus* population in the 1940s at a time of increasing penicillin usage, the acquisition of *mecA* would have provided an additional mechanism of penicillin resistance to the early MRSA. Subsequently, when methicillin was introduced to bypass penicillin resistance, this second generation β-lactam became the selective pressure leading to the successful emergence and spread of the first MRSA lineage.

The appearance and clonal expansion of the first MRSA across Europe was the beginning of an era of increasing resistance in this pathogen associated with the epidemic spread of the health-care adapted MRSA clones. Following the initial success of the archaic MRSA clone, the lineage faded, being replaced in the following decades by multiple other HA-MRSA lineages. The events leading to archaic MRSA clones disappearance are not understood. Its successors such as ST239, carried larger SCC*mec* elements with multiple resistance determinants (Chambers and DeLeo 2009b), so were potentially better adapted to the nosocomial setting in which they flourished. It is interesting therefore to consider if this lineage was restricted in its response to antimicrobials, shaped by a limited range of agents used in the 1940s, and therefore very much adapted to β-lactams. With increased availability and usage of new chemotherapeutic agents, newer lineages emerged able to survive broader spectrum antibiotic prescribing, and consequently were better equipped to survive in comparison to their predecessor in the nosocomial setting.

These events seemingly reflect the unintended consequences of widespread antibiotic usage, and provide a cautionary lesson form history; namely when new antimicrobials are introduced, they may already potentially be ineffective due to previously unrecognised adaptations in the pathogen population. Overall this emphasises the critical importance of population level pathogen surveillance to identify patterns of resistance, and particularly in the context of prescribing practices, and more pertinently during the introduction of new drugs.

7.2 S. aureus *diversity and adaptation during colonisation*

Following on from the population level study of the long-term evolution of *S. aureus,* the work presented in chapters 4 and 5 intended to characterise the micro-evolutionary changes arising in individuals during carriage on the skin. In the first instance the diversity arising during nasal

carriage in healthy children in the community was characterised followed by that occurring in children with moderate to severe AE during disease flares.

The direct comparison of healthy nasal carriers with AE affected children allowed important similarities between the disease groups to be defined. Firstly, *S. aureus* colonisation across the study populations was by clonal populations, expanding and diversifying within the host. Secondly, the observed genetic heterogeneity arising within the colonising populations of each individual was similar, whether colonisation was on nasal epithelium or extra-nasal skin in AE cases. Importantly these findings demonstrated the natural colonisation dynamics of *S. aureus* on the human epithelium.

Whilst there was limited evidence to suggest any adaptation during nasal colonisation in healthy children, specific examples of genetic adaptation were found in AE cases, which were of clinical relevance. This finding is not surprising given the very different clinical contexts of the study populations. Nasal colonisation is asymptomatic, within a comparatively sheltered environment and largely not subjected to any type of physical or therapeutic intervention. Conversely in AE, colonisation is often targeted therapeutically by both antibiotics and antiseptics, and is occurring on body sites continually exposed to environmental influences such as bathing. The combined observations of mutations such as in *fusA* and *agrA*, along with the apparent chronicity of carriage in some AE cases, signals the potential role that intracellular persistence has in *S. aureus* survival in AE. This may be an underappreciated niche for *S. aureus* in AE, and in some individuals is a potential source of recrudescent infection. This could be of particular relevance in the sub-group of children with severe AE disease who are prone to *S. aureus* associated flares whose management necessitates weekly bleach baths and who are typically unresponsive to antibiotics.

The single time point sampling provided only a narrow window through which to assess the colonising populations for diversity and adaptation. Nonetheless, and especially in AE cases it has provided a depth of assessment that has not been previously published. In future, longitudinal assessment of colonisation in AE has significant potential for us to further our understanding of how the organism drives the disease. The time course of colonisation inclusive of pre-, during- and post-flare has not been described in currently published studies. It is important that such studies are conducted, as they would help us investigate whether colonisation triggers a disease flare, and also whether persisting colonisation leads to continued disease activity. Larger scale prospective follow up studies of this sort have the potential, just as they have done in cystic fibrosis, to provide definitive evidence of adaptive mutations. For example, evidence of convergent evolution within and/ or across hosts would provide a route to targeted functional studies that shed more light on the molecular basis of the host-cell interactions in the colonisation of diseased skin.

The impact of therapies on the colonising *S. aureus* population would be another facet that could be investigated by such prospective longitudinal sampling. For instance topical steroids are the mainstay of treatment in AE, and have been shown to reduce *S. aureus* burden on inflamed AE skin with similar, if not better, efficacy than antibiotics (Bath-Hextall et al. 2010; Jensen et al. 2012). Why this is the case is not clear, and has not been studied to date. *In vitro* topical steroid formulations do not have bactericidal activity (Alsterholm et al. 2010). However they have been shown in the context of recurring sinus infection to reduce biofilm formation (Goggin et al. 2014), which suggests they may impact upon intracellular adherence and potentially attachment. Topical antibiotics such as Fusidic acid, which are widely used in AE, have potentially unintended consequences that will promote the survival of *S. aureus,* which include driving persistence as SCVs or in intracellular niches. In addition to serial swab sampling, targeting and characterising intracellular *S. aureus* is also a route to greater understanding of the host-cell interactions of *S. aureus* in AE. Sampling epidermal

keratinocytes with tape stripping, such as is already performed in other aspects of AE research (Kezic et al. 2011), at the same time as swabbing could allow correlation between bacterial genotype and phenotype directly from AE cases.

7.3 Differences in S. aureus *populations carried by AE cases and nasal controls*

Having assessed the colonising populations of nasal carriers and AE affected children from a micro-epidemiological perspective the work presented in chapter 6 aimed to compare isolates from AE cases and nasal controls from a macro-epidemiological perspective.

Assessment of the population structure of isolates from the case and control populations presented in Chapters 4, 5 and 6 respectively, there was evident differences in the clonal backgrounds preferentially colonising AE skin versus the nasal epithelium. In AE, CC1 isolates were more frequent, and in nasal carriage CC30 and 45 predominated. These findings of two geographically distinct studies suggest a segregation of strain on the basis of disease status. Previous studies have concluded that there is no association with AE and preferential strain colonisation (Kim et al. 2009; Yeung et al. 2011; Kong et al. 2012; Rojo et al. 2014), however these studies base their conclusions on small sample sizes. In a Korean AE study CC1 prevalence was detected in AE cases (Kim et al. 2009), and a further study has suggested an association with disease severity in individuals colonised by CC1 isolates (Yeung et al. 2011). A general paucity of CC30 isolates in AE colonisation was also reported by the aforementioned studies.

The high prevalence of CC30 and 45 isolates in nasal carriage studies does of itself suggest that these lineages possess properties making them suited to survival in the mucosal niche. In comparison to nasal epithelium, inflamed AE skin is a vastly differing environment for the organism to survive in, which could be one reason we see CC1 isolates with greater frequency in AE cases. Therapies may also have the potential to select for specific strains with the

preferential ability to colonise the skin in AE, for instance in an individual using topical therapies with benzalkonium chloride, a strain that acquires an antiseptic resistance determinant such as *qacA* would be equipped to survive. Basic functional assessment comparing the strain backgrounds may yield some explanation as to why these differences are observed. For instance given that the cutaneous environments, nasal vs. AE skin, will vary physiologically with regards to temperature and pH, a comparison of growth rates of isolates from differing clonal backgrounds under variation in these parameters would be one obvious experimental probe.

The gene content of the isolate populations were compared to look for differences on the basis of disease state, by multiple methods. From this there were examples of differences in genes, and gene classes that could be relevant for colonisation in AE. However, these findings were largely explained from the underlying clonal influence of the sample collection, which was another limitation of this study in addition to small sample size. More importantly, it serves to highlight the limitations of this approach when trying to make a disease association. As an alternative strategy the pangenome of the isolates were compared using DAPC. Unfortunately, this also failed to separate genetic content of the isolates on the basis of disease. This approach remains a potentially powerful tool to apply in future studies of larger sample collections from AE. In this instance, the small sample size was a significant confounder, alongside issues with assembly quality that were also probably contributory. As a further potential investigative approach, a GWAS (Genome Wide Association study) approach could be applied to datasets such as this. For instance, the recently described sequence element enrichment analysis (SEER) method (Lees et al. 2016). This technique utilises variable *k-mer* length in genome assemblies to allow identification of variable loci of across the genome of a bacterial population, and specifically where there is a polyclonal population background (Lees et al. 2016). This could aid identification of genetic variants, not just gene presence or absence that have an association with disease.

8 References

1000 Genomes Project Consortium, Abecasis, GR, Altshuler, D, *et al.* (2010). A map of human genome variation from population-scale sequencing. *Nature* 467: 1061–73.

Aanensen, DM, Feil, EJ, Holden, MTG, *et al.* (2016). Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe. *MBio* 7: e00444–16.

Abraham, EP, Chain, E (1940). An Enzyme from Bacteria able to Destroy Penicillin. *Nature* 146: 837.

Alamro, M, Bidmos, FA, Chan, H, *et al.* (2014). Phase variation mediates reductions in expression of surface proteins during persistent meningococcal carriage. *Infect Immun* 82: 2472–84.

Allakhverdi, Z, Comeau, MR, Jessup, HK, *et al.* (2007). Thymic stromal lymphopoietin is released by human epithelial cells in response to microbes, trauma, or inflammation and potently activates mast cells. *J Exp Med* 204: 253–8.

Alsterholm, M, Karami, N, Faergemann, J (2010). Antimicrobial activity of topical skin pharmaceuticals - an in vitro study. *Acta Derm Venereol* 90: 239–45.

Altman, D (2000). *Statistics with Confidence: Confidence Intervals and Statistical Guidelines, 2nd Edition*. 2nd edn.

Altschul, SF, Gish, W, Miller, W, *et al.* (1990). Basic local alignment search tool. *J Mol Biol* 215: 403–10.

Aly, R, Maibach, HI, Shinefield, HR (1977). Microbial flora of atopic dermatitis. *Arch Dermatol* 113: 780–2.

Amagai, M, Matsuyoshi, N, Wang, ZH, *et al.* (2000). Toxin in bullous impetigo and staphylococcal scalded-skin syndrome targets desmoglein 1 1–3.

Ardern-Jones, MR, Black, AP, Bateman, EA, *et al.* (2007). Bacterial superantigen facilitates epithelial presentation of allergen to T helper 2 cells. *Proc Natl Acad Sci USA* 104: 5557–62.

Arkwright, PD, Cookson, BD, Haeney, MR, *et al.* (2001). Children with atopic dermatitis who carry toxin-positive Staphylococcus aureus strains have an expansion of blood CD5- B lymphocytes without an increase in disease severity. *Clin Exp Immunol* 125: 184–9.

Aubry-Damon, H, Soussy, CJ, Courvalin, P (1998). Characterization of mutations in the rpoB gene that confer rifampin resistance in Staphylococcus aureus. *Antimicrob Agents Chemother* 42: 2590–4.

Azarian, T, Daum, RS, Petty, LA, *et al.* (2016). Intrahost Evolution of Methicillin-Resistant Staphylococcus aureus USA300 Among Individuals With Reoccurring Skin and Soft-Tissue Infections. *J Infect Dis* 214: 895–905.

Baba, T, Takeuchi, F, Kuroda, M, *et al.* (2002). Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* 359: 1819–27.

Balma-Mena, A, Lara-Corrales, I, Zeller, J, *et al.* (2011). Colonization with community-acquired methicillin-resistant Staphylococcus aureus in children with atopic dermatitis: a cross-sectional study. *Int J Dermatol* 50: 682–8.

Balwit, JM, van Langevelde, P, Vann, JM, *et al.* (1994). Gentamicin-resistant menadione and hemin auxotrophic Staphylococcus aureus persist within cultured endothelial cells. *J Infect Dis* 170: 1033–7.

Barber, M, Rozwadowska-Dowzenko, M (1948). Infection by penicillin-resistant staphylococci. *Lancet* 2: 641–4.

Barbier, F, Ruppé, E, Hernandez, D, *et al.* (2010). Methicillin-resistant coagulase-negative staphylococci in the community: high homology of SCCmec IVa between Staphylococcus epidermidis and major clones of methicillin-resistant Staphylococcus aureus. *J Infect Dis* 202: 270–81.

Basic-Hammer, N, Vogel, V, Basset, P, *et al.* (2010). Impact of recombination on genetic variability within Staphylococcus aureus clonal complexes. *Infect Genet Evol* 10: 1117–23.

Basketter, DA, Marriott, M, Gilmour, NJ, *et al.* (2004). Strong irritants masquerading as skin allergens: the case of benzalkonium chloride. *Contact Derm* 50: 213–7.

Bath-Hextall, FJ, Birnie, AJ, Ravenscroft, JC, *et al.* (2010). Interventions to reduce Staphylococcus aureus in the management of atopic eczema: an updated Cochrane review. *Br J Dermatol* 163: 12–26.

Beck, LA, Boguniewicz, M, Hata, T, *et al.* (2009). Phenotype of atopic dermatitis subjects with a history of eczema herpeticum. *J Allergy Clin Immunol* 124: 260–9–269.e1–7.

Bentley, DR, Balasubramanian, S, Swerdlow, HP, *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53–9.

Berents, TL, Carlsen, KCL, Mowinckel, P, *et al.* (2015). Skin Barrier Function and Staphylococcus aureus Colonization in Vestibulum Nasi and Fauces in Healthy Infants and Infants with Eczema: A Population-Based Cohort Study. *PLoS ONE* 10: e0130145.

Betley, MJ, Mekalanos, JJ (1985). Staphylococcal enterotoxin A is encoded by phage. *Science* 229: 185–7.

Bisacchi, GS, Manchester, JI (2014). A New-Class Antibacterial—Almost. Lessons in Drug Discovery and Development: A Critical Analysis of More than 50 Years of Effort toward ATPase Inhibitors of DNA Gyrase and Topoisomerase IV. *ACS Infect Dis* 1: 4–41.

Bjorland, J, Steinum, T, Sunde, M, *et al.* (2003). Novel plasmid-borne gene qacJ mediates resistance to quaternary ammonium compounds in equine Staphylococcus aureus, Staphylococcus simulans, and Staphylococcus intermedius. *Antimicrob Agents Chemother* 47: 3046–52.

Blumental, S, Deplano, A, Jourdain, S, *et al.* (2013). Dynamic pattern and genotypic diversity of Staphylococcus aureus nasopharyngeal carriage in healthy pre-school children. *J Antimicrob Chemother* 68: 1517–23.

BNF for Children (2016). *British National Formulary for Children*. BMJ Group and Pharmaceutical Press.

Bondi, A, Dietz, CC (1945). Penicillin resistant staphylococci. *Proc Soc Exp Biol Med* 60: 55–8.

Boniface, K, Bernard, F-X, Garcia, M, *et al.* (2005). IL-22 inhibits epidermal differentiation and induces proinflammatory gene expression and migration of human keratinocytes. *J Immunol* 174: 3695–702.

Brauweiler, AM, Bin, L, Kim, BE, *et al.* (2013). Filaggrin-dependent secretion of sphingomyelinase protects against staphylococcal α-toxin-induced keratinocyte death. *J Allergy Clin Immunol* 131: 421–7.e1–2.

Brauweiler, AM, Goleva, E, Leung, DYM (2014). Th2 cytokines increase Staphylococcus aureus alpha toxin-induced keratinocyte death through the signal transducer and activator of transcription 6 (STAT6). *J Invest Dermatol* 134: 2114–21.

Breuer, K, Wittmann, M, Kempe, K, *et al.* (2005). Alpha-toxin is produced by skin colonizing Staphylococcus aureus and induces a T helper type 1 response in atopic dermatitis. *Clin Exp Allergy* 35: 1088–95.

Brown, SJ, McLean, WHI (2012). One remarkable molecule: filaggrin. *J Invest Dermatol* 132: 751–62.

Bunikowski, R, Mielke, ME, Skarabis, H, *et al.* (2000). Evidence for a disease-promoting effect of Staphylococcus aureus-derived exotoxins in atopic dermatitis. *J Allergy Clin Immunol* 105: 814–9.

Bur, S, Preissner, KT, Herrmann, M, *et al.* (2013). The Staphylococcus aureus extracellular adherence protein promotes bacterial internalization by keratinocytes independent of fibronectin-binding proteins. *J Invest Dermatol* 133: 2004–12.

Capoluongo, E, Giglio A, A, Lavieri, MM, *et al.* (2001). Genotypic and phenotypic characterization of Staphylococcus aureus strains isolated in subjects with atopic dermatitis. Higher prevalence of exfoliative B toxin production in lesional strains and correlation between the markers of disease intensity and colonization density. *J Dermatol Sci* 26: 145–55.

Carver, TJ, Rutherford, KM, Berriman, M, *et al.* (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* 21: 3422–3.

Casadevall, A, Pirofski, L-A (2003). The damage-response framework of microbial pathogenesis. *Nat Rev Microbiol* 1: 17–24.

Castanheira, M, Watters, AA, Mendes, RE, *et al.* (2010). Occurrence and molecular characterization of fusidic acid resistance mechanisms among Staphylococcus spp. from European countries (2008). *Journal of Antimicrobial Chemotherapy* 65: 1353–8.

Castillo-Ramírez, S, Corander, J, Marttinen, P, *et al.* (2012). Phylogeographic variation in recombination rates within a global clone of methicillin-resistant Staphylococcus aureus. *Genome Biol* 13: R126.

Castillo-Ramírez, S, Harris, SR, Holden, MTG, *et al.* (2011). The impact of recombination on dN/dS within recently emerged bacterial clones. *PLoS Pathog* 7: e1002129.

Catchpole, I, Thomas, C, Davies, A, *et al.* (1988). The nucleotide sequence of Staphylococcus aureus plasmid pT48 conferring inducible macrolide-lincosamide-streptogramin B resistance and comparison with similar plasmids expressing constitutive resistance. *J Gen Microbiol* 134: 697–709.

CDC (1999). From the Centers for Disease Control and Prevention. Four pediatric deaths from community-acquired methicillin-resistant Staphylococcus aureus--Minnesota and North Dakota, 1997-1999. *JAMA* 282: 1123–5.

Chain, E, Florey, HW, Gardner, AD, *et al.* (1940). PENICILLIN AS A CHEMOTHERAPEUTIC AGENT. *The Lancet* 236: 226–8.

Chambers, HF, DeLeo, FR (2009a). Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat Rev Microbiol* 7: 629–41.

Chambers, HF, DeLeo, FR (2009b). Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat Rev Microbiol* 7: 629–41.

Chang, S, Sievert, DM, Hageman, JC, *et al.* (2003). Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene. *N Engl J Med* 348: 1342–7.

Chen, C-J, Lin, M-H, Shu, J-C, *et al.* (2014). Reduced susceptibility to vancomycin in isogenic Staphylococcus aureus strains of sequence type 59: tracking evolution and identifying mutations by whole-genome sequencing. *J Antimicrob Chemother* 69: 349–54.

Chen, H-J, Hung, W-C, Tseng, S-P, *et al.* (2010). Fusidic acid resistance determinants in Staphylococcus aureus clinical isolates. *Antimicrob Agents Chemother* 54: 4985–91.

Chen, J, Ram, G, Yoong, P, *et al.* (2015). An rpsL-based allelic exchange vector for Staphylococcus aureus. *Plasmid* 79: 8–14.

Chen, L, Yang, J, Yu, J, *et al.* (2005). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 33: D325–8.

Cheung, AL, Eberhardt, KJ, Chung, E, *et al.* (1994). Diminished virulence of a sar-/agr- mutant of Staphylococcus aureus in the rabbit model of endocarditis. *J Clin Invest* 94: 1815–22.

Cheung, GYC, Wang, R, Khan, BA, *et al.* (2011). Role of the accessory gene regulator agr in community-associated methicillin-resistant Staphylococcus aureus pathogenesis. *Infect Immun* 79: 1927–35.

Cho, SH, Strickland, I, Boguniewicz, M, *et al.* (2001a). Fibronectin and fibrinogen contribute to the enhanced binding of Staphylococcus aureus to atopic skin. *J Allergy Clin Immunol* 108: 269–74.

Cho, SH, Strickland, I, Tomkinson, A, *et al.* (2001b). Preferential binding of Staphylococcus aureus to skin sites of Th2-mediated inflammation in a murine model. *J Invest Dermatol* 116: 658–63.

Chuang, Y-Y, Huang, Y-C (2013). Molecular epidemiology of community-associated meticillin-resistant Staphylococcus aureus in Asia. *Lancet Infect Dis* 13: 698–708.

Chung, H-J, Jeon, H-S, Sung, H, *et al.* (2008). Epidemiological characteristics of methicillin-resistant Staphylococcus aureus isolates from children with eczematous atopic dermatitis lesions. *J Clin Microbiol* 46: 991–5.

Colville, JM, Gale, HH, Cox, F, *et al.* (1957). Clinical observations on the use of novobiocin in penicillin-resistant staphylococcal septicemia. *Antibiot Annu* 5: 920–6.

Conceição, T, Aires-de-Sousa, M, Füzi, M, *et al.* (2007). Replacement of methicillin-resistant Staphylococcus aureus clones in Hungary over time: a 10-year surveillance study. *Clin*

Microbiol Infect 13: 971–9.
Corrigan, RM, Miajlovic, H, Foster, TJ (2009). Surface proteins that promote adherence of Staphylococcus aureus to human desquamated nasal epithelial cells. *BMC Microbiol* 9: 22.
Corrigan, RM, Rigby, D, Handley, P, *et al.* (2007). The role of Staphylococcus aureus surface protein SasG in adherence and biofilm formation. *Microbiology (Reading, Engl)* 153: 2435–46.
Couto, I, de Lencastre, H, Severina, E, *et al.* (1996). Ubiquitous presence of a mecA homologue in natural isolates of Staphylococcus sciuri. *Microb Drug Resist* 2: 377–91.
Crisóstomo, MI, Westh, H, Tomasz, A, *et al.* (2001). The evolution of methicillin resistance in Staphylococcus aureus: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. *Proc Natl Acad Sci USA* 98: 9865–70.
Croucher, NJ, Harris, SR, Fraser, C, *et al.* (2011). Rapid pneumococcal evolution in response to clinical interventions. *Science* 331: 430–4.
Crump, B, Wise, R, Dent, J (1983). Pharmacokinetics and tissue penetration of ciprofloxacin. *Antimicrob Agents Chemother* 24: 784–6.
Das, S, Lindemann, C, Young, BC, *et al.* (2016). Natural mutations in a Staphylococcus aureus virulence regulator attenuate cytotoxicity but permit bacteremia and abscess formation. *Proc Natl Acad Sci USA* 113: E3101–10.
Dastgheyb, SS, Otto, M (2015). Staphylococcal adaptation to diverse physiologic niches: an overview of transcriptomic and phenotypic changes in different biological environments. *Future Microbiol* 10: 1981–95.
DeLeo, FR, Otto, M, Kreiswirth, BN, *et al.* (2010). Community-associated meticillin-resistant Staphylococcus aureus. *Lancet* 375: 1557–68.
Demerec, M (1949). Patterns of bacterial resistance to Penicillin, Aureomycin, and Streptomycin. *J Clin Invest* 28: 891–3.
Dickinson, TM, Archer, GL (2000). Phenotypic expression of oxacillin resistance in Staphylococcus epidermidis: roles of mecA transcriptional regulation and resistant-subpopulation selection. *Antimicrob Agents Chemother* 44: 1616–23.
Didelot, X, Walker, AS, Peto, TE, *et al.* (2016). Within-host evolution of bacterial pathogens. *Nat Rev Microbiol* 14: 150–62.
Diep, BA, Gill, SR, Chang, RF, *et al.* (2006). Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus. *Lancet* 367: 731–9.
Domagk, G (1936). Chemotherapie der Streptokokken-Infektionen. *Klin Wochenschr* 15: 1585–90.
Dordel, J, Kim, C, Chung, M, *et al.* (2014). Novel determinants of antibiotic resistance: identification of mutated loci in highly methicillin-resistant subpopulations of methicillin-resistant Staphylococcus aureus. *MBio* 5: e01000–e01000–13.
Dreyfus, DH (2013). Herpesviruses and the microbiome. *J Allergy Clin Immunol* 132: 1278–86.
Drummond, AJ, Suchard, MA, Xie, D, *et al.* (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29: 1969–73.
Duggar, BM (1948). Aureomycin; a product of the continuing search for new antibiotics. *Ann N Y Acad Sci* 51: 177–81.
Dyke, KG, Jevons, MP, Parker, MT (1966). Penicillinase production and intrinsic resistance to penicillins in Staphylococcus aures. *Lancet* 1: 835–8.
Eichenfield, LF, Tom, WL, Berger, TG, *et al.* (2014). Guidelines of care for the management of atopic dermatitis. *J Am Acad Dermatol* 71: 116–32.
Eiff, von, C, Becker, K, Metze, D, *et al.* (2001). Intracellular persistence of Staphylococcus aureus small-colony variants within keratinocytes: a cause for antibiotic treatment failure in a patient with darier's disease. *Clin Infect Dis* 32: 1643–7.
Elias, PM (2005). Stratum Corneum Defensive Functions: An Integrated View. *J Invest Dermatol* 125: 183–200.
Elias, PM, Wakefield, JS (2014). Mechanisms of abnormal lamellar body secretion and the

dysfunctional skin barrier in patients with atopic dermatitis. *J Allergy Clin Immunol* 134: 781–1.

Ellington, MJ, Hope, R, Livermore, DM, *et al.* (2010). Decline of EMRSA-16 amongst methicillin-resistant Staphylococcus aureus causing bacteraemias in the UK between 2001 and 2007. *J Antimicrob Chemother* 65: 446–8.

Enright, MC, Day, NP, Davies, CE, *et al.* (2000). Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. *J Clin Microbiol* 38: 1008–15.

Enright, MC, Robinson, DA, Randle, G, *et al.* (2002). The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). *Proc Natl Acad Sci USA* 99: 7687–92.

Eriksen, KR, Erichsen, I (1964). Resistance to methicillin, isoxazolyl penicillins, and cephalothin in Staphylococcus aureus. *Acta Pathol Microbiol Scand* 62: 255–75.

Esposito, S, Terranova, L, Zampiero, A, *et al.* (2014). Oropharyngeal and nasal Staphylococcus aureus carriage by healthy children. *BMC Infect Dis* 14: 723.

Evans, RJ, Waterworth, PM (1966). Naturally-occurring fusidic acid resistance in staphylococci and its linkage to other resistances. *Journal of Clinical Pathology* 19: 555–60.

Everitt, RG, Didelot, X, Batty, EM, *et al.* (2014). Mobile elements drive recombination hotspots in the core genome of Staphylococcus aureus. *Nat Commun* 5: 3956.

Ewing, B, Hillier, L, Wendl, MC, *et al.* (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175–85.

Fairbrother, RW, Williams, BL (1956). Two new antibiotics; antibacterial activity of novobiocin and vancomycin. *Lancet* 271: 1177–8.

Feil, EJ, Cooper, JE, Grundmann, H, *et al.* (2003). How clonal is Staphylococcus aureus? *J Bacteriol* 185: 3307–16.

Feil, EJ, Maiden, MC, Achtman, M, *et al.* (1999). The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis. *Mol Biol Evol* 16: 1496–502.

Fitzgerald, JR (2014). Evolution of Staphylococcus aureus during human colonization and infection. *Infect Genet Evol* 21: 542–7.

Fitzgerald, JR, Holden, MTG (2016). Genomics of Natural Populations of Staphylococcus aureus. *Annu Rev Microbiol* 70: 459–78.

Fitzgerald, JR, Sturdevant, DE, Mackie, SM, *et al.* (2001). Evolutionary genomics of Staphylococcus aureus: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc Natl Acad Sci USA* 98: 8821–6.

Fleming, A (1945). Penicillin. *Nobel Lecture* 1–11.

Foster, TJ, Geoghegan, JA, Ganesh, VK, *et al.* (2014). Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. *Nat Rev Microbiol* 12: 49–62.

Fowler, VG, Sakoulas, G, McIntyre, LM, *et al.* (2004). Persistent bacteremia due to methicillin-resistant Staphylococcus aureus infection is associated with agr dysfunction and low-level in vitro resistance to thrombin-induced platelet microbicidal protein. *J Infect Dis* 190: 1140–9.

Franke, AE, Clewell, DB (1985). EvidenceforaChromosome-BorneResistanceTransposon (Tn916)inStreptococcusfaecalisThatIsCapableof "Conjugal"TransferintheAbsenceofaConjugativePlasmid. *J Bacteriol* 145: 494–502.

Furi, L, Ciusa, ML, Knight, D, *et al.* (2013). Evaluation of reduced susceptibility to quaternary ammonium compounds and bisbiguanides in clinical isolates and laboratory-generated mutants of Staphylococcus aureus. *Antimicrob Agents Chemother* 57: 3488–97.

Gao, W, Chua, K, Davies, JK, *et al.* (2010). Two novel point mutations in clinical Staphylococcus aureus reduce linezolid susceptibility and switch on the stringent response to promote persistent infection. *PLoS Pathog* 6: e1000944.

García-Álvarez, L, Holden, MTG, Lindsay, H, *et al.* (2011). Meticillin-resistant Staphylococcus aureus with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study. *Lancet Infect Dis* 11: 595–603.

Gill, SR, Fouts, DE, Archer, GL, *et al.* (2005). Insights on evolution of virulence and resistance

from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain. *J Bacteriol* 187: 2426–38.

Gittler, JK, Shemer, A, Suárez-Fariñas, M, *et al.* (2012). Progressive activation of T(H)2/T(H)22 cytokines and selective epidermal proteins characterizes acute and chronic atopic dermatitis. *J Allergy Clin Immunol* 130: 1344–54.

Goggin, R, Jardeleza, C, Wormald, P-J, *et al.* (2014). Corticosteroids directly reduce Staphylococcus aureus biofilm growth: an in vitro study. *Laryngoscope* 124: 602–7.

Goh, CL, Wong, JS, Giam, YC (1997). Skin colonization of Staphylococcus aureus in atopic dermatitis patients seen at the National Skin Centre, Singapore. *Int J Dermatol* 36: 653–7.

Golubchik, T, Batty, EM, Miller, RR, *et al.* (2013). Within-host evolution of Staphylococcus aureus during asymptomatic carriage. *PLoS ONE* 8: e61319.

Gomes, AR, Westh, H, de Lencastre, H (2006). Origins and evolution of methicillin-resistant Staphylococcus aureus clonal lineages. *Antimicrob Agents Chemother* 50: 3237–44.

Grandgirard, D, Furi, L, Ciusa, ML, *et al.* (2015). Mutations upstream of fabI in triclosan resistant Staphylococcus aureus strains are associated with elevated fabI gene expression. *BMC Genomics* 16: 345.

Grice, EA, Kong, HH, Conlan, S, *et al.* (2009). Topographical and temporal diversity of the human skin microbiome. *Science* 324: 1190–2.

Grice, EA, Segre, JA (2011). The skin microbiome. *Nat Rev Microbiol* 9: 244–53.

Griffith, LJ (1966). Development of resistance to kanamycin. *Ann N Y Acad Sci* 132: 796–9.

Griggs, DJ, Marona, H, Piddock, LJV (2003). Selection of moxifloxacin-resistant Staphylococcus aureus compared with five other fluoroquinolones. *J Antimicrob Chemother* 51: 1403–7.

Grumann, D, Nübel, U, Bröker, BM (2014). Staphylococcus aureus toxins--their functions and genetics. *Infect Genet Evol* 21: 583–92.

Guang-Sen, S, Boost, M, Cho, P (2016). Prevalence of antiseptic resistance genes increases in staphylococcal isolates from orthokeratology lens wearers over initial six-month period of use. *Eur J Clin Microbiol Infect Dis* 35: 955–62.

Guzik, TJ, Bzowska, M, Kasprowicz, A, *et al.* (2005). Persistent skin colonization with Staphylococcus aureus in atopic dermatitis: relationship to clinical and immunological parameters. *Clin Exp Allergy* 35: 448–55.

Hanifin, JM, Thurston, M, Omoto, M, *et al.* (2001). The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol* 10: 11–8.

Hansson, S, Singh, R, Gudkov, AT, *et al.* (2005). Structural insights into fusidic acid resistance and sensitivity in EF-G. *J Mol Biol* 348: 939–49.

Harris, SR, Cartwright, EJP, Török, ME, *et al.* (2013). Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis* 13: 130–6.

Harris, SR, Clarke, IN, Seth-Smith, HMB, *et al.* (2012). Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet* 44: 413–9–S1.

Harris, SR, Feil, EJ, Holden, MTG, *et al.* (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327: 469–74.

Harris, SR, Robinson, C, Steward, KF, *et al.* (2015). Genome specialization and decay of the strangles pathogen, Streptococcus equi, is driven by persistent infection. *Genome Res* 25: 1360–71.

Hartman, BJ, Tomasz, A (1984). Low-affinity penicillin-binding protein associated with beta-lactam resistance in Staphylococcus aureus. *J Bacteriol* 158: 513–6.

Hauk, PJ, Hamid, QA, Chrousos, GP, *et al.* (2000). Induction of corticosteroid insensitivity in human PBMCs by microbial superantigens. *J Allergy Clin Immunol* 105: 782–7.

Hauser, C, Wuethrich, B, Matter, L, *et al.* (1985). Staphylococcus aureus colonization in atopic dermatitis patients. *Dermatologica*.

Hayes, SM, Howlin, R, Johnston, DA, *et al.* (2015). Intracellular residency of Staphylococcus

aureus within mast cells in nasal polyps: A novel observation. *J Allergy Clin Immunol* 135: 1648–51.

Heijer, den, CDJ, van Bijnen, EME, Paget, WJ, *et al.* (2013). Prevalence and resistance of commensal Staphylococcus aureus, including meticillin-resistant S aureus, in nine European countries: a cross-sectional study. *Lancet Infect Dis* 13: 409–15.

Heng, NCK, Wescombe, PA, Burton, JP, *et al.* (2007). The Diversity of Bacteriocins in Gram-Positive Bacteria. *In*: *Bacteriocins*, Bacteriocins: Ecology and Evolution. Springer Berlin Heidelberg: Berlin, Heidelberg, 45–92.

Hiramatsu, K, Cui, L, Kuroda, M, *et al.* (2001). The emergence and evolution of methicillin-resistant Staphylococcus aureus 9: 486–93.

Hiramatsu, K, Hanaki, H, Ino, T, *et al.* (1997). Methicillin-resistant Staphylococcus aureus clinical strain with reduced vancomycin susceptibility. *J Antimicrob Chemother* 40: 135–6.

Hirasawa, Y, Takai, T, Nakamura, T, *et al.* (2010). Staphylococcus aureus extracellular protease causes epidermal barrier dysfunction. *J Invest Dermatol* 130: 614–7.

Hoeger, PH, Lenz, W, Boutonnier, A, *et al.* (1992). Staphylococcal Skin Colonization in Children with Atopic Dermatitis: Prevalence, Persistence, and Transmission. *J Infect Dis* 165: 1064–8.

Holden, MTG, Feil, EJ, Lindsay, JA, *et al.* (2004). Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci USA* 101: 9786–91.

Holden, MTG, Hsu, LY, Kurt, K, *et al.* (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Res* 23: 653–64.

Holden, MTG, Lindsay, JA, Corton, C, *et al.* (2010). Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant Staphylococcus aureus, sequence type 239 (TW). *J Bacteriol* 192: 888–92.

Holtfreter, S, Grumann, D, Schmudde, M, *et al.* (2007). Clonal distribution of superantigen genes in clinical Staphylococcus aureus isolates. *J Clin Microbiol* 45: 2669–80.

Hood, MI, Skaar, EP (2012). Nutritional immunity: transition metals at the pathogen–host interface. *Nat Rev Microbiol* 10: 525–37.

Horsburgh, MJ, Wharton, SJ, Cox, AG, *et al.* (2002). MntR modulates expression of the PerR regulon and superoxide resistance in Staphylococcus aureus through control of manganese uptake. *Mol Microbiol* 44: 1269–86.

Howden, BP, Grayson, LM (2006). Dumb and Dumber—The Potential Waste of a Useful Antistaphylococcal Agent: Emerging Fusidic Acid Resistance in Staphylococcus aureus. *Clinical Infectious Diseases* 394–400.

Howden, BP, McEvoy, CRE, Allen, DL, *et al.* (2011). Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR. *PLoS Pathog* 7: e1002359.

Howell, MD, Boguniewicz, M, Pastore, S, *et al.* (2006). Mechanism of HBD-3 deficiency in atopic dermatitis. *Clin Immunol* 121: 332–8.

Huang, T-W, Chen, F-J, Miu, W-C, *et al.* (2012). Complete genome sequence of Staphylococcus aureus M013, a pvl-positive, ST59-SCCmec type V strain isolated in Taiwan. *J Bacteriol* 194: 1256–7.

Huesca, M, Peralta, R, Sauder, DN, *et al.* (2002). Adhesion and Virulence Properties of Epidemic Canadian Methicillin-Resistant Staphylococcus aureus Strain 1: Identification of Novel Adhesion Functions Associated with Plasmin-Sensitive Surface Protein. *J Infect Dis* 185: 1285–96.

Hurdle, JG, O'Neill, AJ, Mody, L, *et al.* (2005). In vivo transfer of high-level mupirocin resistance from Staphylococcus epidermidis to methicillin-resistant Staphylococcus aureus associated with failure of mupirocin prophylaxis. *J Antimicrob Chemother* 56: 1166–8.

Inouye, M, Dashnow, H, Raven, L-A, *et al.* (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 6: 90.

Ip, M, Wang, Z, Lam, WY, *et al.* (2014). Draft Genome Sequence of Methicillin-Resistant

Staphylococcus aureus CUHK_188 (ST188), a Health Care-Associated Bacteremic Isolate from Hong Kong. *Genome Announc* 2: e00255–14–e00255–14.

Irvine, AD, McLean, WHI, Leung, DYM (2011). Filaggrin mutations associated with skin and allergic diseases. *N Engl J Med* 365: 1315–27.

Ito, T, Katayama, Y, Asada, K, *et al.* (2001). Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant Staphylococcus aureus. *Antimicrob Agents Chemother* 45: 1323–36.

Iwase, T, Uehara, Y, Shinji, H, *et al.* (2010). Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization. *Nature* 465: 346–9.

IWG-SCC (2009). Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. *Antimicrob Agents Chemother* 53: 4961–7.

Jagadeesan, S, Kurien, G, Divakaran, MV, *et al.* (2014). Methicillin-resistant Staphylococcus aureus colonization and disease severity in atopic dermatitis: a cross-sectional study from South India. *Indian J Dermatol Venereol Leprol* 80: 229–34.

Jensen, JM, Scherer, A, Wanke, C, *et al.* (2012). Gene expression is differently affected by pimecrolimus and betamethasone in lesional skin of atopic dermatitis. *Allergy* 67: 423.

Jessen, O, Rosendal, K, Bülow, P, *et al.* (1969). Changing staphylococci and staphylococcal infections. A ten-year study of bacteria and cases of bacteremia. *N Engl J Med* 281: 627–35.

Jevons, MP (1961). “Celbenin-” resistant Staphylococci. *Br Med J* 124–5.

Johnson, AP, Aucken, HM, Cavendish, S, *et al.* (2001). Dominance of EMRSA-15 and -16 among MRSA causing nosocomial bacteraemia in the UK: analysis of isolates from the European Antimicrobial Resistance Surveillance System (EARSS). *J Antimicrob Chemother* 48: 143–4.

Jombart, T (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–5.

Jombart, T, Devillard, S, Balloux, F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11: 94.

Juuti, K, Ibrahem, S, Virolainen-Julkunen, A, *et al.* (2005). The pls gene found in methicillin-resistant Staphylococcus aureus strains is common in clinical isolates of Staphylococcus sciuri. *J Clin Microbiol* 43: 1415–9.

Juuti, KM, Sinha, B, Werbick, C, *et al.* (2004). Reduced adherence and host cell invasion by methicillin-resistant Staphylococcus aureus expressing the surface protein Pls. *J Infect Dis* 189: 1574–84.

Kaatz, GW, Thyagarajan, RV, Seo, SM (2005). Effect of promoter region mutations and mgrA overexpression on transcription of norA, which encodes a Staphylococcus aureus multidrug efflux transporter. *Antimicrob Agents Chemother* 49: 161–9.

Kahl, B, Herrmann, M, Everding, AS, *et al.* (1998). Persistent infection with small colony variant strains of Staphylococcus aureus in patients with cystic fibrosis. *J Infect Dis* 177: 1023–9.

Kaneko, J, Kimura, T, Narita, S, *et al.* (1998). Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Panton-Valentine leukocidin genes. *Gene* 215: 57–67.

Katayama, Y, Ito, T, Hiramatsu, K (2000). A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. *Antimicrob Agents Chemother* 44: 1549–55.

Katayama, Y, Ito, T, Hiramatsu, K (2001). Genetic organization of the chromosome region surrounding mecA in clinical staphylococcal strains: role of IS431-mediated mecI deletion in expression of resistance in mecA-carrying, low-level methicillin-resistant Staphylococcus haemolyticus. *Antimicrob Agents Chemother* 45: 1955–63.

Kehl-Fie, TE, Zhang, Y, Moore, JL, *et al.* (2013). MntABC and MntH contribute to systemic Staphylococcus aureus infection by competing with calprotectin for nutrient manganese. *Infect Immun* 81: 3395–405.

Kennedy, EA, Connolly, J, O'B Hourihane, J, *et al.* (2016). Skin microbiome prior to development of atopic dermatitis: early colonization with commensal staphylococci at 2

months is associated with a lower risk of atopic dermatitis at 1 year. *J Allergy Clin Immunol*.

Kezic, S, O'Regan, GM, Yau, N, *et al.* (2011). Levels of filaggrin degradation products are influenced by both filaggrin genotype and atopic dermatitis severity. *Allergy* 66: 934–40.

Kim, D-W, Park, J-Y, Park, K-D, *et al.* (2009). Are there predominant strains and toxins of Staphylococcus aureus in atopic dermatitis patients? Genotypic characterization and toxin determination of S. aureus isolated in adolescent and adult patients with atopic dermatitis. *J Dermatol* 36: 75–81.

Kim, SA, Kim, BR, Chun, MY, *et al.* (2016). Relation between pH in the Trunk and Face: Truncal pH Can Be Easily Predicted from Facial pH. *Ann Dermatol* 28: 216–21.

Kirby, WM (1944). Extraction of a highly potent penicillin inactivator from penicillin resistant staphylococci. *Science* 99: 452–3.

Kirst, HA, Thompson, DG, Nicas, TI (1998). Historical yearly usage of vancomycin. *Antimicrob Agents Chemother* 42: 1303–4.

Kluytmans, J, van Belkum, A, Verbrugh, H (1997). Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. *Clin Microbiol Rev* 10: 505–20.

Knight, GM, Budd, EL, Whitney, L, *et al.* (2012). Shift in dominant hospital-associated methicillin-resistant Staphylococcus aureus (HA-MRSA) clones over time. *J Antimicrob Chemother* 67: 2514–22.

Knox, R (1960). A new penicillin (BRL 1241) active against penicillin-resistant staphylococci. *Br Med J* 2: 690–3.

Kong, HH, Oh, J, Deming, C, *et al.* (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res* 22: 850–9.

Köser, CU, Holden, MTG, Ellington, MJ, *et al.* (2012). Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 366: 2267–75.

Kreiswirth, B, Kornblum, J, Arbeit, RD, *et al.* (1993). Evidence for a clonal origin of methicillin resistance in Staphylococcus aureus. *Science* 259: 227–30.

Krell, T, Lacal, J, Busch, A, *et al.* (2010). Bacterial sensor kinases: diversity in the recognition of environmental signals. *Annu Rev Microbiol* 64: 539–59.

Kuo, I-H, Yoshida, T, De Benedetto, A, *et al.* (2013). The cutaneous innate immune response in patients with atopic dermatitis. *J Allergy Clin Immunol* 131: 266–78.

Kuroda, M, Ohta, T, Uchiyama, I, *et al.* (2001). Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *Lancet* 357: 1225–40.

Lai, Y, Cogen, AL, Radek, KA, *et al.* (2010). Activation of TLR2 by a small molecule produced by Staphylococcus epidermidis increases antimicrobial defense against bacterial skin infections. *J Invest Dermatol* 130: 2211–21.

Lannergård, J, Eiff, von, C, Sander, G, *et al.* (2008). Identification of the genetic basis for clinical menadione-auxotrophic small-colony variant isolates of Staphylococcus aureus. *Antimicrob Agents Chemother* 52: 4017–22.

Le, KY, Otto, M (2015). Quorum-sensing regulation in staphylococci-an overview. *Front Microbiol* 6: 1174.

Lebon, A, Labout, JAM, Verbrugh, HA, *et al.* (2008). Dynamics and determinants of Staphylococcus aureus carriage in infancy: the Generation R Study. *J Clin Microbiol* 46: 3517–21.

Leclercq, R, Derlot, E, Duval, J, *et al.* (1988). Plasmid-mediated resistance to vancomycin and teicoplanin in Enterococcus faecium. *N Engl J Med* 319: 157–61.

Lees, JA, Vehkala, M, Välimäki, N, *et al.* (2016). Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat Commun* 7: 12797.

Leung, DY (1995). Atopic dermatitis: the skin as a window into the pathogenesis of chronic allergic diseases. *J Allergy Clin Immunol* 96: 302–18–quiz319.

Leung, DYM, Guttman-Yassky, E (2014). Deciphering the complexities of atopic dermatitis: shifting paradigms in treatment approaches. *J Allergy Clin Immunol* 134: 769–79.

Lever, R, Hadley, K, Downey, D, *et al.* (1988). Staphylococcal colonization in atopic dermatitis and the effect of topical mupirocin therapy. *Br J Dermatol* 119: 189–98.

Levine, DP (2006). Vancomycin: a history. *Clin Infect Dis* 42 Suppl 1: S5–12.

Levinson, DC, Griffith, GC, Pearson, HE (1951). Antibiotics in Management of Staphylococcal Endocarditis. *Calif Med* 74: 167–70.

Lewis, K (2013). Platforms for antibiotic discovery. *Nature Publishing Group* 12: 371–87.

Leyden, JJ, Marples, RR, Kligman, AM (1974). Staphylococcus aureus in the lesions of atopic dermatitis. *Br J Dermatol* 90: 525–30.

Li, H, Durbin, R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–60.

Li, H, Handsaker, B, Wysoker, A, *et al.* (2009a). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–9.

Li, M, Diep, BA, Villaruz, AE, *et al.* (2009b). Evolution of virulence in epidemic community-associated methicillin-resistant Staphylococcus aureus. *Proc Natl Acad Sci USA* 106: 5883–8.

Lina, G, Durand, G, Berchich, C, *et al.* (2006). Staphylococcal chromosome cassette evolution in Staphylococcus aureus inferred from ccr gene complex sequence typing analysis. *Clin Microbiol Infect* 12: 1175–84.

Lindsay, JA (2014). Staphylococcus aureus genomics and the impact of horizontal gene transfer. *Int J Med Microbiol* 304: 103–9.

Lindsay, JA, Holden, MTG (2004). Staphylococcus aureus: superbug, super genome? *Trends Microbiol* 12: 378–85.

Lindsay, JA, Holden, MTG (2006). Understanding the rise of the superbug: investigation of the evolution and genomic variation of Staphylococcus aureus. *Funct Integr Genomics* 6: 186–201.

Lindsay, JA, Knight, GM, Budd, EL, *et al.* (2012). Shuffling of mobile genetic elements (MGEs) in successful healthcare-associated MRSA (HA-MRSA). *Mob Genet Elements* 2: 239–43.

Lindsay, JA, Moore, CE, Day, NP, *et al.* (2006). Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes. *J Bacteriol* 188: 669–76.

Lindsay, JA, Ruzin, A, Ross, HF, *et al.* (1998). The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in Staphylococcus aureus. *Mol Microbiol* 29: 527–43.

Littlejohn, TG, DiBerardino, D, Messerotti, LJ, *et al.* (1991). Structure and evolution of a family of genes encoding antiseptic and disinfectant resistance in Staphylococcus aureus. *Gene* 101: 59–66.

Lo, W-T, Wang, S-R, Tseng, M-H, *et al.* (2010). Comparative molecular analysis of meticillin-resistant Staphylococcus aureus isolates from children with atopic dermatitis and healthy subjects in Taiwan. *British Journal of Dermatology* 162: 1110–6.

Lomholt, H, Andersen, KE, Kilian, M (2005). Staphylococcus aureus clonal dynamics and virulence factors in children with atopic dermatitis. *J Invest Dermatol* 125: 977–82.

Lowy, FD (2003). Antimicrobial resistance: the example of Staphylococcus aureus. *J Clin Invest* 111: 1265–73.

Ma, XX, Ito, T, Tiensasitorn, C, *et al.* (2002). Novel type of staphylococcal cassette chromosome mec identified in community-acquired methicillin-resistant Staphylococcus aureus strains. *Antimicrob Agents Chemother* 46: 1147–52.

Maiden, MC, Bygraves, JA, Feil, E, *et al.* (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95: 3140–5.

Malachowa, N, Whitney, AR, Kobayashi, SD, *et al.* (2011). Global changes in Staphylococcus aureus gene expression in human blood. *PLoS ONE* 6: e18617.

Markussen, T, Marvig, RL, Gómez-Lozano, M, *et al.* (2014). Environmental heterogeneity drives within-host diversification and evolution of Pseudomonas aeruginosa. *MBio* 5: e01592–14.

Marvig, RL, Damkiær, S, Khademi, SMH, *et al.* (2014). Within-host evolution of Pseudomonas aeruginosa reveals adaptation toward iron acquisition from hemoglobin. *MBio* 5: e00966–

14.
Marvig, RL, Johansen, HK, Molin, S, *et al.* (2013). Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet* 9: e1003741.
Marvig, RL, Sommer, LM, Molin, S, *et al.* (2015). Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nat Genet* 47: 57–64.
Matsui, K, Nishikawa, A (2003). Lipoteichoic acid from Staphylococcus aureus enhances allergen-specific immunoglobulin E production in mice. *Clin Exp Allergy* 33: 842–8.
Matsui, K, Wirotesangthong, M, Nishikawa, A (2007). Percutaneous application of peptidoglycan from Staphylococcus aureus induces eosinophil infiltration in mouse skin. *Clin Exp Allergy* 37: 615–22.
Matthews, P, Tomasz, A (1990). Insertional inactivation of the mec gene in a transposon mutant of a methicillin-resistant clinical isolate of Staphylococcus aureus. *Antimicrob Agents Chemother* 34: 1777–9.
McAleer, MA, Irvine, AD (2013). The multifunctional role of filaggrin in allergic skin disease. *J Allergy Clin Immunol* 131: 280–91.
McCarthy, AJ, Breathnach, AS, Lindsay, JA (2012). Detection of mobile-genetic-element variation between colonizing and infecting hospital-associated methicillin-resistant Staphylococcus aureus isolates. *J Clin Microbiol* 50: 1073–5.
McCarthy, AJ, Lindsay, JA (2010). Genetic variation in Staphylococcus aureus surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. *BMC Microbiol* 10: 173.
McCarthy, AJ, Lindsay, JA (2012). The distribution of plasmids that carry virulence and resistance genes in Staphylococcus aureus is lineage associated. *BMC Microbiol* 12: 104.
McCarthy, AJ, Lindsay, JA (2013). Staphylococcus aureus innate immune evasion is lineage-specific: a bioinfomatics study. *Infect Genet Evol* 19: 7–14.
McGuire, JM, Bunch, RL, Anderson, RC, *et al.* (1952). Ilotycin, a new antibiotic. *Antibiot Chemother (Northfield)* 2: 281–3.
McKenna, A, Hanna, M, Banks, E, *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–303.
Melles, DC, Gorkink, RFJ, Boelens, HAM, *et al.* (2004). Natural population dynamics and expansion of pathogenic clones of Staphylococcus aureus. *J Clin Invest* 114: 1732–40.
Miajlovic, H, Fallon, PG, Irvine, AD, *et al.* (2010). Effect of filaggrin breakdown products on growth of and protein expression by Staphylococcus aureus. *J Allergy Clin Immunol* 126: 1184–90.e3.
Milheiriço, C, Oliveira, DC, de Lencastre, H (2007). Multiplex PCR strategy for subtyping the staphylococcal cassette chromosome mec type IV in methicillin-resistant Staphylococcus aureus: 'SCCmec IV multiplex'. *J Antimicrob Chemother* 60: 42–8.
Miller, SJ, Aly, R, Shinefeld, HR, *et al.* (1988). In vitro and in vivo antistaphylococcal activity of human stratum corneum lipids. *Arch Dermatol* 124: 209–15.
Mitra, A, Mohanraj, M, Shah, M (2009). High levels of fusidic acid-resistant Staphylococcus aureus despite restrictions on antibiotic use. *Clin Exp Dermatol* 34: 136–9.
Monecke, S, Coombs, G, Shore, AC, *et al.* (2011). A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant Staphylococcus aureus. *PLoS ONE* 6: e17936.
Monecke, S, Luedicke, C, Slickers, P, *et al.* (2009). Molecular epidemiology of Staphylococcus aureus in asymptomatic carriers. *Eur J Clin Microbiol Infect Dis* 28: 1159–65.
Monecke, S, Müller, E, Büchler, J, *et al.* (2014a). Staphylococcus aureus in vitro secretion of alpha toxin (hla) correlates with the affiliation to clonal complexes. *PLoS ONE* 9: e100427.
Monecke, S, Müller, E, Büchler, J, *et al.* (2014b). Staphylococcus aureus In Vitro Secretion of Alpha Toxin (hla) Correlates with the Affiliation to Clonal Complexes. *PLoS ONE* 9: e100427–8.
Montgomery, CP, Boyle-Vavra, S, Daum, RS (2010). Importance of the global regulators Agr and SaeRS in the pathogenesis of CA-MRSA USA300 infection. *PLoS ONE* 5: e15177.

Moore, PC, Lindsay, JA (2001). Genetic variation among hospital isolates of methicillin-sensitive Staphylococcus aureus: evidence for horizontal transfer of virulence genes. *J Clin Microbiol* 39: 2760–7.

Moxon, R, Bayliss, C, Hood, D (2006). Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* 40: 307–33.

Mulcahy, ME, Geoghegan, JA, Monk, IR, *et al.* (2012). Nasal colonisation by Staphylococcus aureus depends upon clumping factor B binding to the squamous epithelial cell envelope protein loricrin. *PLoS Pathog* 8: e1003092.

Mulligan, ME, Ruane, PJ, Johnston, L, *et al.* (1987). Ciprofloxacin for eradication of methicillin-resistant Staphylococcus aureus colonization. *Am J Med* 82: 215–9.

Murphy, E, Huwyler, L, de Freire Bastos, MDC (1985). Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. *EMBO J* 4: 3357–65.

Mwangi, MM, Wu, SW, Zhou, Y, *et al.* (2007). Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. *Proc Natl Acad Sci USA* 104: 9451–6.

Na, S-Y, Roh, J-Y, Kim, J-M, *et al.* (2012). Analysis of Colonization and Genotyping of the Exotoxins of Staphylococcus aureus in Patients with Atopic Dermatitis. *Ann Dermatol* 24: 413.

Nakamura, Y, Oscherwitz, J, Cease, KB, *et al.* (2013). Staphylococcus δ-toxin induces allergic skin disease by activating mast cells. *Nature* 503: 397–401.

National Institute for Health and Care Excellence (2007). Atopic eczema in under 12s: diagnosis and management 1–41.

Newman, RL, Bhat, KM, Hackney, R, *et al.* (1962). Fusidic acid: laboratory and clinical assessment. *Br Med J* 2: 1645–7.

Noall, EW, Sewards, HF, Waterworth, PM (1962). Successful Treatment of a Case of Proteus Septicaemia. *Br Med J* 2: 1101–2.

Norström, T, Lannergård, J, Hughes, D (2007). Genetic and phenotypic identification of fusidic acid-resistant mutants with the small-colony-variant phenotype in Staphylococcus aureus. *Antimicrob Agents Chemother* 51: 4438–46.

Noto, MJ, Archer, GL (2006). A subset of Staphylococcus aureus strains harboring staphylococcal cassette chromosome mec (SCCmec) type IV is deficient in CcrAB-mediated SCCmec excision. *Antimicrob Agents Chemother* 50: 2782–8.

Novick, RP, Projan, SJ, Kornblum, J, *et al.* (1995). The agr P2 operon: an autocatalytic sensory transduction system in Staphylococcus aureus. *Mol Gen Genet* 248: 446–58.

Nübel, U, Strommenger, B, Layer, F, *et al.* (2011). From types to trees: reconstructing the spatial spread of Staphylococcus aureus based on DNA variation. *Int J Med Microbiol* 301: 614–8.

O'Neill, AJ, Larsen, AR, Henriksen, AS, *et al.* (2004). A fusidic acid-resistant epidemic strain of Staphylococcus aureus carries the fusB determinant, whereas fusA mutations are prevalent in other resistant isolates. *Antimicrob Agents Chemother* 48: 3594–7.

O'Neill, AJ, McLaws, F, Kahlmeter, G, *et al.* (2007). Genetic basis of resistance to fusidic acid in staphylococci. *Antimicrob Agents Chemother* 51: 1737–40.

Okuma, K, Iwakawa, K, Turnidge, JD, *et al.* (2002). Dissemination of new methicillin-resistant Staphylococcus aureus clones in the community. *J Clin Microbiol* 40: 4289–94.

Ong, PY, Leung, DYM (2010). The infectious aspects of atopic dermatitis. *Immunol Allergy Clin North Am* 30: 309–21.

Ong, PY, Ohtake, T, Brandt, C, *et al.* (2002). Endogenous antimicrobial peptides and skin infection in Atopic Dermatitis. *N Engl J Med* 347: 1151–60.

Oranje, AP, Glazenburg, EJ, Wolkerstorfer, A, *et al.* (2007). Practical issues on interpretation of scoring atopic dermatitis: the SCORAD index, objective SCORAD and the three-item severity score. *Br J Dermatol* 157: 645–8.

Otto, M (2010). Staphylococcus colonization of the skin and antimicrobial peptides. *Expert Rev Dermatol* 5: 183–95.

Otto, M (2013). Coagulase-negative staphylococci as reservoirs of genes facilitating MRSA infection: Staphylococcal commensal species such as Staphylococcus epidermidis are being recognized as important sources of genes promoting MRSA colonization and virulence. *Bioessays* 35: 4–11.

Otto, M (2014). Staphylococcus aureus toxins. *Curr Opin Microbiol* 17: 32–7.

Page, AJ, Cummins, CA, Hunt, M, *et al.* (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* btv421.

Page, AJ, Taylor, B, Delaney, AJ, *et al.* (2016). SNP-sites: rapid efficient extraction of SNPs from multi- FASTA alignments. *Microbial Genomics* 1–5.

Palmer, CNA, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38: 441–6.

Park, J-M, Jo, J-H, Jin, H, *et al.* (2016). Change in Antimicrobial Susceptibility of Skin-Colonizing Staphylococcus aureus in Korean Patients with Atopic Dermatitis during Ten-Year Period. *Ann Dermatol* 28: 470–8.

Pascolini, C, Sinagra, J, Pecetta, S, *et al.* (2011). Molecular and immunological characterization of Staphylococcus aureus in pediatric atopic dermatitis: implications for prophylaxis and clinical management. *Clin Dev Immunol* 2011: 718708–7.

Paterson, GK, Harrison, EM, Murray, GGR, *et al.* (2015). Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat Commun* 6: 6560.

Paterson, GK, Larsen, AR, Robb, A, *et al.* (2012). The newly described mecA homologue, mecALGA251, is present in methicillin-resistant Staphylococcus aureus isolates from a diverse range of host species. *J Antimicrob Chemother* 67: 2809–13.

Paulander, W, Nissen Varming, A, Bæk, KT, *et al.* (2013). Antibiotic-mediated selection of quorum-sensing-negative Staphylococcus aureus. *MBio* 3: e00459–12.

Peleg, AY, Miyakis, S, Ward, DV, *et al.* (2012). Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of Staphylococcus aureus. *PLoS ONE* 7: e28316.

Peterson, LR, Quick, JN, Jensen, B, *et al.* (1990). Emergence of ciprofloxacin resistance in nosocomial methicillin-resistant Staphylococcus aureus isolates. Resistance during ciprofloxacin plus rifampin therapy for methicillin-resistant S aureus colonization. *Arch Intern Med (Chic)* 150: 2151–5.

Peton, V, Bouchard, DS, Almeida, S, *et al.* (2014). Fine-tuned characterization of Staphylococcus aureus Newbould 305, a strain associated with mild and chronic mastitis in bovines. *Vet Res* 45: 106.

Phillips, S, Novick, RP (1979). Tn554--a site-specific repressor-controlled transposon in Staphylococcus aureus. *Nature* 278: 476–8.

Price, EP, Sarovich, DS, Mayo, M, *et al.* (2013). Within-host evolution of Burkholderia pseudomallei over a twelve-year chronic carriage infection. *MBio* 4: e00388–13–e00388–13.

Price, LB, Stegger, M, Hasman, H, *et al.* (2012). Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock. *MBio* 3: e00305–11–e00305–11.

Proctor, RA, Eiff, von, C, Kahl, BC, *et al.* (2006a). Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat Rev Microbiol* 4: 295–305.

Proctor, RA, Eiff, von, C, Kahl, BC, *et al.* (2006b). Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat Rev Microbiol* 4: 295–305.

Public Health England (2015). Bacteraemia: Voluntary surveillance of bacteraemia in England, Wales and Northern Ireland- 2007 to 2014. *Health Protection Report* 9: 1–12.

Pyle, MM (1947). Relative numbers of resistant tubercle bacilli in sputa of patients before and during treatment with streptomycin. *Proc Staff Meet Mayo Clin* 22: 465–73.

Rambaut, A, Lam, TT, Max Carvalho, L, *et al.* (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2: vew007.

Ravenscroft, JC, Layton, AM, Eady, EA, *et al.* (2003). Short-term effects of topical fusidic acid or

mupirocin on the prevalence of fusidic acid resistant (FusR) Staphylococcus aureus in atopic eczema. *Br J Dermatol* 148: 1010–7.

Reginald, K, Westritschnig, K, Linhart, B, *et al.* (2011). Staphylococcus aureus fibronectin-binding protein specifically binds IgE from patients with atopic dermatitis and requires antigen presentation for cellular immune responses. *J Allergy Clin Immunol* 128: 82–8.

Rhem, MN, Lech, EM, Patti, JM, *et al.* (2000). The collagen-binding adhesin is a virulence factor in Staphylococcus aureus keratitis. *Infect Immun* 68: 3776–9.

Richardson, JF, Reith, S (1993). Characterization of a strain of methicillin-resistant Staphylococcus aureus (EMRSA-15) by conventional and molecular methods. *J Hosp Infect* 25: 45–52.

Riley, M, Space, DB (1996). Genes and proteins of Escherichia coli (GenProtEc). *Nucleic Acids Res* 24: 40.

Rippke, F, Schreiner, V, Doering, T, *et al.* (2004). Stratum corneum pH in atopic dermatitis: impact on skin barrier function and colonization with Staphylococcus Aureus. *Am J Clin Dermatol* 5: 217–23.

Roantree, RJ, Rantz, LA (1955). Fatal staphylococcal endocarditis treated with erythromycin; rapid development of resistance in vivo and in vitro. *AMA Arch Intern Med* 95: 320–5.

Roberts, RB, Tennenberg, AM, Eisner, W, *et al.* (1998). Outbreak in a New York City teaching hospital burn center caused by the Iberian epidemic clone of MRSA. *Microb Drug Resist* 4: 175–83.

Robinson, DA, Monk, AB, Cooper, JE, *et al.* (2005). Evolutionary genetics of the accessory gene regulator (agr) locus in Staphylococcus aureus. *J Bacteriol* 187: 8312–21.

Roche, FM (2003). Characterization of novel LPXTG-containing proteins of Staphylococcus aureus identified from genome sequences. *Microbiology (Reading, Engl)* 149: 643–54.

Rodríguez, EA, Correa, MM, Ospina, S, *et al.* (2014). Differences in epidemiological and molecular characteristics of nasal colonization with Staphylococcus aureus (MSSA-MRSA) in children from a university hospital and day care centers. *PLoS ONE* 9: e101417.

Roduit, C, Frei, R, Loss, G, *et al.* (2012). Development of atopic dermatitis according to age of onset and association with early-life exposures. *J Allergy Clin Immunol* 130: 130–5.

Rojo, A, Aguinaga, A, Monecke, S, *et al.* (2014). Staphylococcus aureus genomic pattern and atopic dermatitis: may factors other than superantigens be involved? *Eur J Clin Microbiol Infect Dis* 33: 651–8.

Rolinson, GN, Stevens, S, Batchelor, FR, *et al.* (1960). Bacteriological studies on a new penicillin-BRL. 1241. *Lancet* 2: 564–7.

Rountree, PM (1955). Infections caused by a particular phage type of Staphylococcus aureus. 42: 157–61.

Rowland, SJ, Dyke, KG (1989). Characterization of the staphylococcal beta-lactamase transposon Tn552. *EMBO J* 8: 2761–73.

Rutherford, K, Parkhill, J, Crook, J, *et al.* (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–5.

Salah, LA, Faergemann, J (2015). A retrospective analysis of skin bacterial colonisation, susceptibility and resistance in atopic dermatitis and impetigo patients. *Acta Derm Venereol* 95: 532–5.

Schlievert, PM, Case, LC, Strandberg, KL, *et al.* (2008). Superantigen profile of Staphylococcus aureus isolates from patients with steroid-resistant atopic dermatitis. *Clin Infect Dis* 46: 1562–7.

Schlievert, PM, Strandberg, KL, Lin, Y-C, *et al.* (2010). Secreted virulence factor comparison between methicillin-resistant and methicillin-sensitive Staphylococcus aureus, and its relevance to atopic dermatitis. *J Allergy Clin Immunol* 125: 39–49.

Schmitt, J, Spuls, PI, Thomas, KS, *et al.* (2014). The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. *J Allergy Clin Immunol* 134: 800–7.

Schnellmann, C, Gerber, V, Rossano, A, *et al.* (2006). Presence of new mecA and mph(C) variants conferring antibiotic resistance in Staphylococcus spp. isolated from the skin of

horses before and after clinic admission. *J Clin Microbiol* 44: 4444–54.
Schoenbach, EB, Bryer, MS, Long, PH (1948). The pharmacology and clinical trial of Aureomycin: A preliminary report. *Ann N Y Acad Sci* 51: 267–79.
Seemann, T (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30: 2068–9.
Sendi, P, Proctor, RA (2009). Staphylococcus aureus as an intracellular pathogen: the role of small colony variants. *Trends Microbiol* 17: 54–8.
Serres, MH, Gopal, S, Nahum, LA, *et al.* (2001). A functional update of the Escherichia coli K-12 genome. *Genome Biol* 2: RESEARCH0035.
Shamssain, M (2007). Trends in the prevalence and severity of asthma, rhinitis, and atopic eczema in 6- to 7- and 13- to 14- yr old children from the north-east of England. *Pediatr Allergy Immunol* 18: 149.
Shopsin, B, Drlica-Wagner, A, Mathema, B, *et al.* (2008). Prevalence of agr dysfunction among colonizing Staphylococcus aureus strains. *J Infect Dis* 198: 1171–4.
Shopsin, B, Eaton, C, Wasserman, GA, *et al.* (2010). Mutations in agr do not persist in natural populations of methicillin-resistant Staphylococcus aureus. *J Infect Dis* 202: 1593–9.
Shore, A, Rossney, AS, Keane, CT, *et al.* (2005). Seven Novel Variants of the Staphylococcal Chromosomal Cassette mec in Methicillin-Resistant Staphylococcus aureus Isolates from Ireland. *Antimicrob Agents Chemother* 49: 2070–83.
Silver, LL (2011). Challenges of antibacterial discovery. *Clin Microbiol Rev* 24: 71–109.
Skinner, D, Keefer, C (1941). Significance of bacteremia caused by Staphylococcus aureus. *Arch Intern Med (Chic)* 68: 851–75.
Smith, EE, Buckley, DG, Wu, Z, *et al.* (2006). Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc Natl Acad Sci USA* 103: 8487–92.
Smith, EM, Green, LE, Medley, GF, *et al.* (2005). Multilocus sequence typing of intercontinental bovine Staphylococcus aureus isolates. *J Clin Microbiol* 43: 4737–43.
Smyth, DS, Feil, EJ, Meaney, WJ, *et al.* (2009). Molecular genetic typing reveals further insights into the diversity of animal-associated Staphylococcus aureus. *J Med Microbiol* 58: 1343–53.
Sollid, JUE, Furberg, AS, Hanssen, AM, *et al.* (2014). Staphylococcus aureus: determinants of human carriage. *Infect Genet Evol* 21: 531–41.
Somerville, GA, Proctor, RA (2009). At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci. *Microbiol Mol Biol Rev* 73: 233–48.
Sonkoly, E, Muller, A, Lauerma, AI, *et al.* (2006). IL-31: a new link between T cells and pruritus in atopic skin inflammation. *J Allergy Clin Immunol* 117: 411–7.
Soong, G, Paulino, F, Wachtel, S, *et al.* (2015). Methicillin-resistant Staphylococcus aureus adaptation to human keratinocytes. *MBio* 6: e00289–15.
Spaulding, AR, Salgado-Pabón, W, Kohler, PL, *et al.* (2013). Staphylococcal and streptococcal superantigen exotoxins. *Clin Microbiol Rev* 26: 422–47.
Spink, WW, Ferris, V (1945). Quantitative action of penicillin inhibitor from penicillin-resistant strains of Staphylococci. *Science* 102: 221–3.
Stamatakis, A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–90.
Stanczak-Mrozek, KI, Manne, A, Knight, GM, *et al.* (2015). Within-host diversity of MRSA antimicrobial resistances. *J Antimicrob Chemother* 70: 2191–8.
Stegger, M, Driebe, EM, Roe, C, *et al.* (2013). Genome Sequence of Staphylococcus aureus Strain CA-347, a USA600 Methicillin-Resistant Isolate. *Genome Announc* 1: e00517–13–e00517–13.
Stewart, GT, Nixon, HH, Coles, HM, *et al.* (1960). Report on clinical use of BRL 1241 in children with staphylococcal and streptococcal infections. *Br Med J* 2: 703–6.
Stieger, M, Angehrn, P, Wohlgensinger, B, *et al.* (1996). GyrB mutations in Staphylococcus aureus strains resistant to cyclothialidine, coumermycin, and novobiocin. *Antimicrob Agents Chemother* 40: 1060–2.
Strange, P, Skov, L, Lisby, S, *et al.* (1996). Staphylococcal enterotoxin B applied on intact normal and intact atopic skin induces dermatitis. *Arch Dermatol* 132: 27–33.

Suerbaum, S, Smith, JM, Bapumia, K, *et al.* (1998). Free recombination within Helicobacter pylori. *Proc Natl Acad Sci USA* 95: 12619–24.

Sung, JM-L, Lloyd, DH, Lindsay, JA (2008). Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology (Reading, Engl)* 154: 1949–59.

Tauber, M, Balica, S, Hsu, C-Y, *et al.* (2016). Staphylococcus aureus density on lesional and nonlesional skin is strongly associated with disease severity in atopic dermatitis. *J Allergy Clin Immunol* 137: 1272–3.

Terada, M, Tsutsui, H, Imai, Y, *et al.* (2006). Contribution of IL-18 to atopic-dermatitis-like skin inflammation induced by Staphylococcus aureus product in mice. *Proc Natl Acad Sci USA* 103: 8816–21.

Thammavongsa, V, Kern, JW, Missiakas, DM, *et al.* (2009). Staphylococcus aureus synthesizes adenosine to escape host immune responses. *J Exp Med* 206: 2417–27.

Thammavongsa, V, Kim, HK, Missiakas, D, *et al.* (2015). Staphylococcal manipulation of host immune responses. *Nat Rev Microbiol* 13: 529–43.

Thurlow, LR, Joshi, GS, Clark, JR, *et al.* (2013). Functional modularity of the arginine catabolic mobile element contributes to the success of USA300 methicillin-resistant Staphylococcus aureus. *Cell Host Microbe* 13: 100–7.

Tong, SYC, Holden, MTG, Nickerson, EK, *et al.* (2015). Genome sequencing defines phylogeny and spread of methicillin-resistant Staphylococcus aureus in a high transmission setting. *Genome Res* 25: 111–8.

Totté, JEE, van der Feltz, WT, Hennekam, M, *et al.* (2016). Prevalence and odds of Staphylococcus aureus carriage in atopic dermatitis: a systematic review and meta-analysis. *British Journal of Dermatology* 175: 687–95.

Traber, KE, Lee, E, Benson, S, *et al.* (2008). agr function in clinical Staphylococcus aureus isolates. *Microbiology (Reading, Engl)* 154: 2265–74.

Travers, JB, Kozman, A, Mousdicas, N, *et al.* (2010). Infected atopic dermatitis lesions contain pharmacologic amounts of lipoteichoic acid. *J Allergy Clin Immunol* 125: 146–52.e1–2.

Trzcinski, K, Cooper, BS, Hryniewicz, W, *et al.* (2000). Expression of resistance to tetracyclines in strains of methicillin-resistant Staphylococcus aureus. *J Antimicrob Chemother* 45: 763–70.

Tuchscherr, L, Medina, E, Hussain, M, *et al.* (2011). Staphylococcus aureus phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. *EMBO Mol Med* 3: 129–41.

Udo, EE, Pearman, JW, Grubb, WB (1993). Genetic analysis of community isolates of methicillin-resistant Staphylococcus aureus in Western Australia. *J Hosp Infect* 25: 97–108.

Uhlemann, A-C, Dordel, J, Knox, JR, *et al.* (2014a). Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. *Proc Natl Acad Sci USA* 111: 6738–43.

Uhlemann, A-C, Otto, M, Lowy, FD, *et al.* (2014b). Evolution of community- and healthcare-associated methicillin-resistant Staphylococcus aureus. *Infect Genet Evol* 21: 563–74.

Umezawa, H, Ueda, M, Maeda, K, *et al.* (1957). Production and isolation of a new antibiotic: kanamycin. *J Antibiot* 10: 181–8.

van Belkum, A, Melles, DC, Nouwen, J, *et al.* (2009a). Co-evolutionary aspects of human colonisation and infection by Staphylococcus aureus. *Infect Genet Evol* 9: 32–47.

van Belkum, A, Verkaik, NJ, de Vogel, CP, *et al.* (2009b). Reclassification of Staphylococcus aureus nasal carriage types. *J Infect Dis* 199: 1820–6.

van den Akker, ELT, Nouwen, JL, Melles, DC, *et al.* (2006). Staphylococcus aureus nasal carriage is associated with glucocorticoid receptor gene polymorphisms. *J Infect Dis* 194: 814–8.

van Hal, SJ, Steen, JA, Espedido, BA, *et al.* (2014). In vivo evolution of antimicrobial resistance in a series of Staphylococcus aureus patient isolates: the entire picture or a cautionary tale? *J Antimicrob Chemother* 69: 363–7.

van Wamel, WJB, Rooijakkers, SHM, Ruyken, M, *et al.* (2006). The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus

aureus are located on beta-hemolysin-converting bacteriophages. *J Bacteriol* 188: 1310–5.
Viana, D, Comos, M, McAdam, PR, *et al.* (2015). A single natural nucleotide mutation alters bacterial pathogen host tropism. *Nat Genet* 47: 361–6.
Vickers, AA, Potter, NJ, Fishwick, CWG, *et al.* (2009). Analysis of mutational resistance to trimethoprim in Staphylococcus aureus by genetic and structural modelling techniques. *J Antimicrob Chemother* 63: 1112–7.
Votintseva, AA, Miller, RR, Fung, R, *et al.* (2014). Multiple-strain colonization in nasal carriers of Staphylococcus aureus. *J Clin Microbiol* 52: 1192–200.
Walsh, EJ, O'Brien, LM, Liang, X, *et al.* (2004). Clumping factor B, a fibrinogen-binding MSCRAMM (microbial surface components recognizing adhesive matrix molecules) adhesin of Staphylococcus aureus, also binds to the tail region of type I cytokeratin 10. *J Biol Chem* 279: 50691–9.
Wang, L, Archer, GL (2010). Roles of CcrA and CcrB in excision and integration of staphylococcal cassette chromosome mec, a Staphylococcus aureus genomic island. *J Bacteriol* 192: 3204–12.
Warner, JA, McGirt, LY, Beck, LA (2009). Biomarkers of Th2 polarity are predictive of staphylococcal colonization in subjects with atopic dermatitis. *Br J Dermatol* 160: 183–5.
Watanabe, S, Ito, T, Sasaki, T, *et al.* (2009). Genetic diversity of staphylocoagulase genes (coa): insight into the evolution of variable chromosomal virulence factors in Staphylococcus aureus. *PLoS ONE* 4: e5714.
Weinert, LA, Chaudhuri, RR, Wang, J, *et al.* (2015). Genomic signatures of human and animal disease in the zoonotic pathogen Streptococcus suis. *Nat Commun* 6: 6740.
Wertheim, HFL, Melles, DC, Vos, MC, *et al.* (2005). The role of nasal carriage in Staphylococcus aureus infections. *Lancet Infect Dis* 5: 751–62.
Willems, RJL, Hanage, WP, Bessen, DE, *et al.* (2011). Population biology of Gram-positive pathogens: high-risk clones for dissemination of antibiotic resistance. *FEMS Microbiol Rev* 35: 872–900.
Williams, RE, Gibson, AG, Aitchison, TC, *et al.* (1990). Assessment of a contact-plate sampling technique and subsequent quantitative bacterial studies in atopic dermatitis. *Br J Dermatol* 123: 493–501.
Wisplinghoff, H, Rosato, AE, Enright, MC, *et al.* (2003). Related clones containing SCCmec type IV predominate among clinically significant Staphylococcus epidermidis isolates. *Antimicrob Agents Chemother* 47: 3574–9.
Woodward, TE, Smadel, JE (1948). Preliminary report on the beneficial effect of chloromycetin in the treatment of typhoid fever. *Ann Intern Med* 29: 131–4.
Wyllie, D, Paul, J, Crook, D (2011). Waves of trouble: MRSA strain dynamics and assessment of the impact of infection control. *J Antimicrob Chemother* 66: 2685–8.
Yang, Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–91.
Yarwood, JM, McCormick, JK, Schlievert, PM (2001). Identification of a novel two-component regulatory system that acts in global regulation of virulence factors of Staphylococcus aureus. *J Bacteriol* 183: 1113–23.
Ye, K, Schulz, MH, Long, Q, *et al.* (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25: 2865–71.
Yeung, M, Balma-Mena, A, Shear, N, *et al.* (2011). Identification of major clonal complexes and toxin producing strains among Staphylococcus aureus associated with atopic dermatitis. *Microbes Infect* 13: 189–97.
Young, BC, Golubchik, T, Batty, EM, *et al.* (2012). Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. *Proc Natl Acad Sci USA* 109: 4550–5.
Yuki, T, Yoshida, H, Akazawa, Y, *et al.* (2011). Activation of TLR2 enhances tight junction barrier in epidermal keratinocytes. *J Immunol* 187: 3230–7.
Zerbino, DR, Birney, E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821–9.

Zhang, K, Sparling, J, Chow, BL, *et al.* (2004). New quadriplex PCR assay for detection of methicillin and mupirocin resistance and simultaneous discrimination of Staphylococcus aureus from coagulase-negative staphylococci. *J Clin Microbiol* 42: 4947–55.

Zhang, Y, Young, D (1994). Molecular genetics of drug resistance in Mycobacterium tuberculosis. *J Antimicrob Chemother* 34: 313–9.

Zheng, T, Yu, J, Oh, MH, *et al.* (2011). The atopic march: progression from atopic dermatitis to allergic rhinitis and asthma. *Allergy Asthma Immunol Res* 3: 67–73.

Zollner, TM, Wichelhaus, TA, Hartung, A, *et al.* (2000). Colonization with superantigen-producing Staphylococcus aureus is associated with increased severity of atopic dermatitis. *Clin Exp Allergy* 30: 994–1000.

Appendix A: Chapter 4 Supplementary data

Staphylococcus aureus carriage in healthy children

The following results are the characterisation of the remaining 15 nasal controls presented in Chapter 4. For each individuals sequenced colonies is a representative phylogeny, and where identified a table detailing the SNPs and indels arising within that controls samples.

Control study participant SS_016

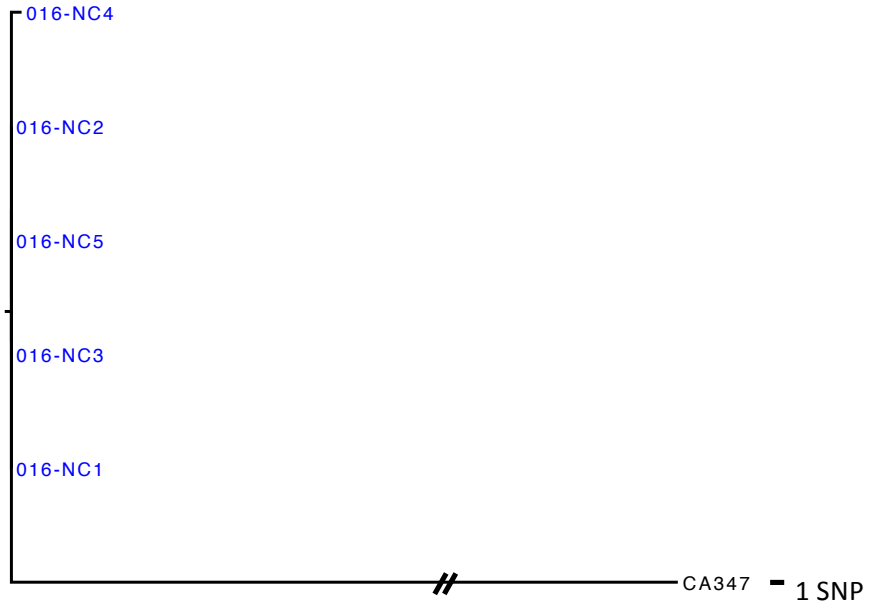

Figure S1 Maximum likelihood core SNPs tree for control study ID SS_016. Tree of 5 sequenced nasal colonies. Colonisation was by ST45 strain. Tree was rooted ST45 reference (CA347). SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-1 In host-heterogeneity identified in nasal carriage control Study ID SS_016. SNPs differentiating colonies sequenced colonies. Base position and region/ gene are relation to position in ST45 reference (CA347; accession number: CP006044). NS– non-synonymous.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
952810	T->A	NS	CA347_895 *clpB* ATP-dependent chaperone protein ClpB	L->I	NC4

No unique insertions or deletions were identified in SS_016.

Control study participant SS_026

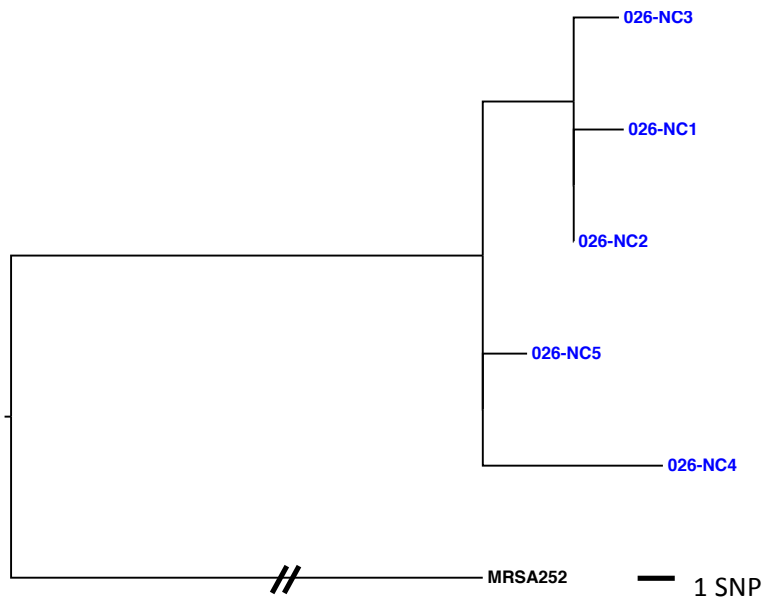

Figure S-2 Maximum likelihood core SNPs tree for control study ID SS_026. Tree of 5 sequenced nasal colonies. Colonisation was by ST30 strain. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-2 In host-heterogeneity identified in nasal carriage control Study ID SS_026. SNPs differentiating sequenced colonies. Base position and region/ gene are relation to position in MRSA252 reference (accession number: BX571856). NS– non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
823767	A->G	NS	SAR0785- ribonucleoside-diphosphate reductase alpha chain	K->R	NC4
1878541	G->A	NS	SAR1803 PTS system IIBC component	T->I	NC4
1591073	C->A	NS	SAR1495 putative lipoprotein	V-> L	NC1
926781	C->T	NS	SAR0888 conserved hypothetical protein	S->L	NC3
2022462	G->A	NS	SAR1940 putative histidine kinase	T->M	NC5
1484563	A->T	NS	SAR1425 *citK* 2-oxoglutarate dehydrogenase E1 component	V->D	NC2/1/3
1267143	A->G	S	SAR1216 *trmD* putative tRNA (guanine-7-)-methyltransferase		NC4
1230776	C->T	I	Between SAR1183 and SAR1184		NC4
988458	A->G	I	Between SAR0944 and SAR0945		NC2/1/3

No unique insertions or deletions were identified in SS_026.

Control study participant SS_045

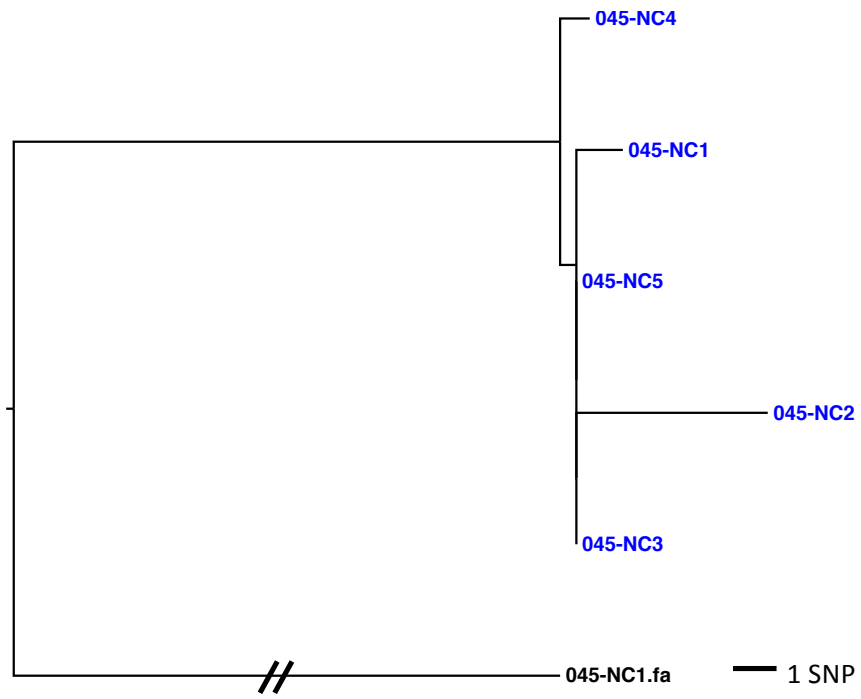

Figure S3 Maximum likelihood core SNPs tree for control study ID SS_045. Tree of 5 sequenced nasal colonies. Colonisation was by ST15 strain. Tree was rooted using self-assembly of nose colony 1 (045-NC1.fa). SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-3-1 Within host diversity identified in nasal carriage control Study ID SS_045. Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region are in relation to self-colony used for mapping (045-NC1.fa) with annotation with annotation transferred from MSSA476. NS– non-synonymous; S-synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
2293886	G->A	NS	SAS1558 putative cysteine desulfurase	A->V	NC2
2433919	A->T	NS	SAS1676 putative D-alanine aminotransferase	I->N	NC2
50802	T->C	NS	SAS2537 conserved hypothetical protein	V->A	NC2/3/4/5
2586622	T->C	NS	SAS2332 putative membrane protein	F->L	NC4
289481	T->G	I	Between SAS0229 (hypothetical protein) and SAS0230a (hypothetical protein)		NC2

Table S-3-2 Within host diversity identified in nasal carriage control Study ID SS_045. Unique indels in the core genome differentiating sequenced colonies. Base position and region are in relation to self-colony used for mapping (045-NC1.fa) with annotation with annotation transferred from MSSA476. (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
289492	NC2	(+) A	Intergenic region between SAR2390 (putative monooxygenase) and SAR2391 (putative N-acetylmuramoyl-L-alanine amidase)	Intergenic; No consequence
2623743	NC3	(-) cat	present in SAR2496- putative solute binding lipoprotein	Deletion of single AA in histidine repeat region of H; no predicted consequence

Control study participant SS_057

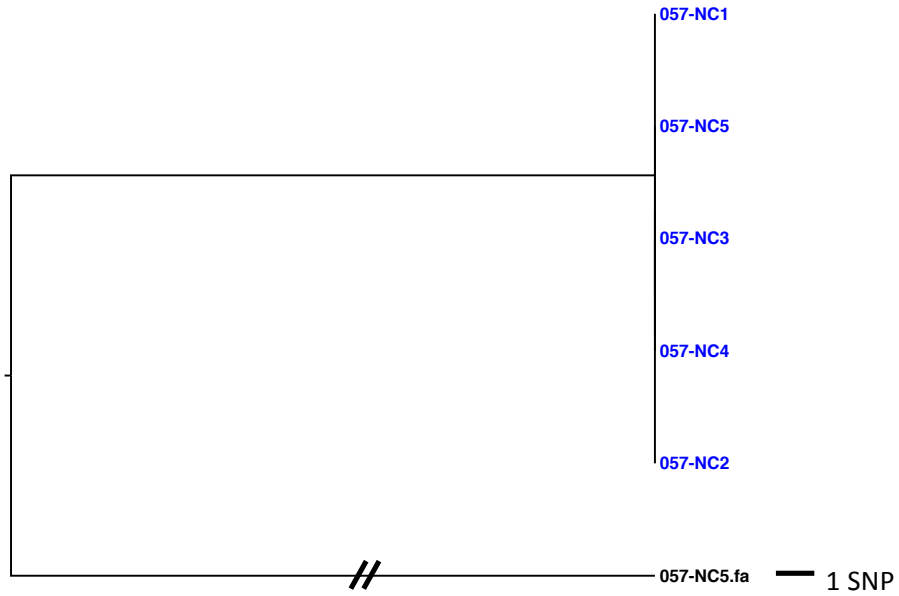

Figure S-4 Maximum likelihood core SNPs tree for control study ID SS_057. Tree of 5 sequenced nasal colonies. Colonisation was by ST582 strain. All colonies are genetically identical at core genome level. Tree was rooted using self-assembly of nose colony 5 (057-NC5.fa). SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

No unique SNPs, insertions or deletions were identified in SS_057.

Control study participant SS_059

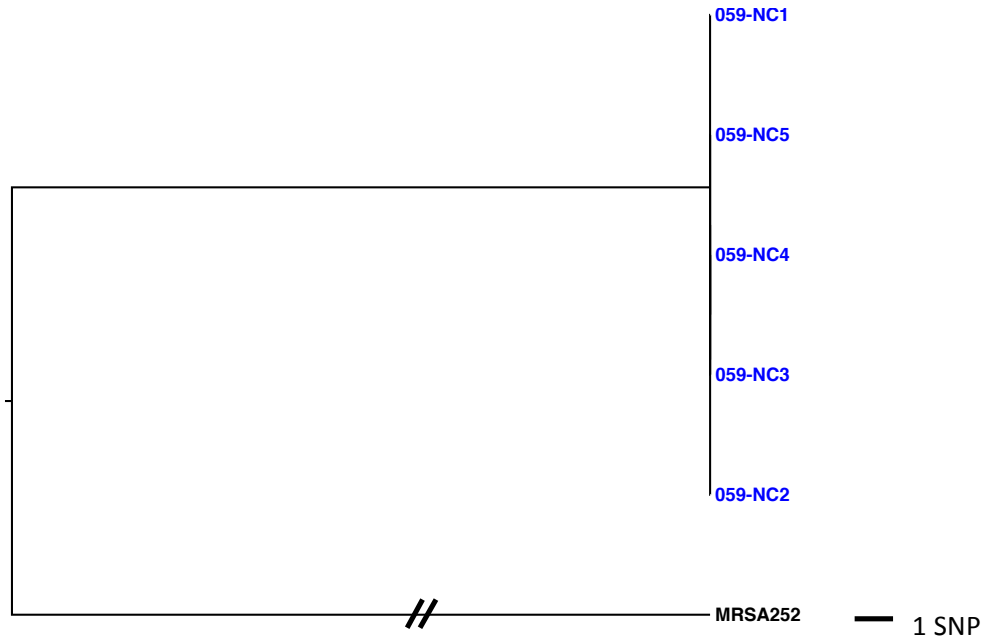

Figure S-5 Maximum likelihood core SNPs tree for control study ID SS_059. Tree of 5 sequenced nasal colonies. Colonisation was by ST30 strain. All colonies are genetically identical at core genome level. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

No unique SNPs, insertions or deletions were identified in SS_059.

Control study participant SS_091

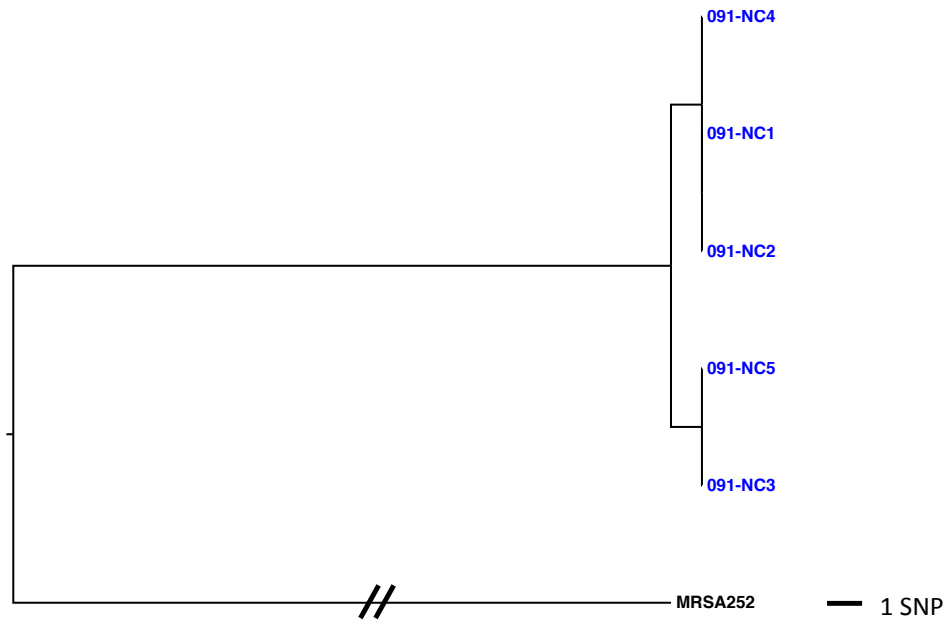

Figure S-6 Maximum likelihood core SNPs tree for control study ID SS_091. Tree of 5 sequenced nasal colonies. Colonisation was by ST30 strain. All colonies are genetically identical at core genome level. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-4 Within host-heterogeneity identified in nasal carriage control Study ID SS_091. Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in MRSA252 reference (accession number: BX571856). S- synonymous.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
1089465	C->T	S	SAR1044 *purF* putative amidophosphoribosyltransferase precursor		NC1/2/4

No unique insertions or deletions were identified in SS_091.

Control study participant SS_094

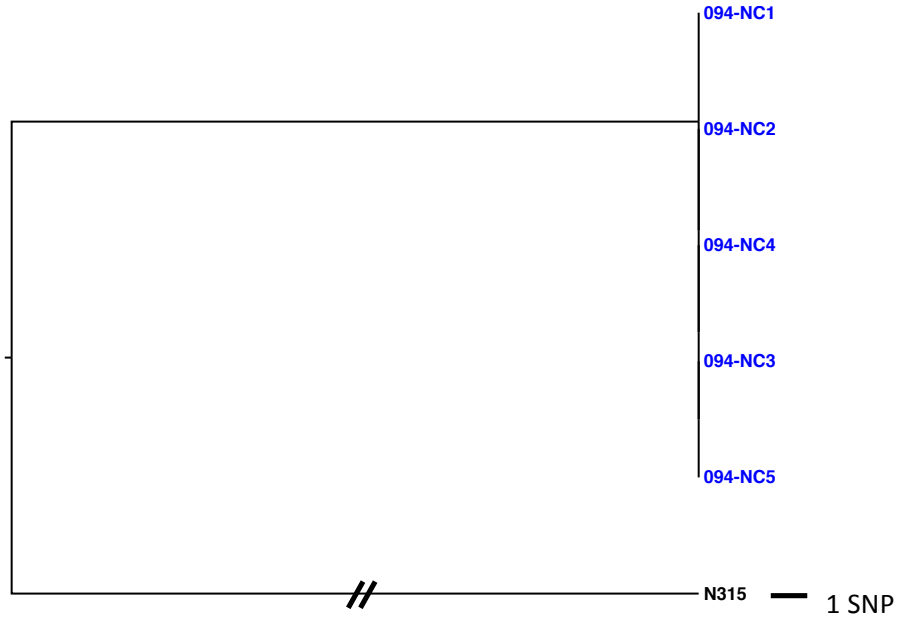

Figure S-7 Maximum likelihood core SNPs tree for control study ID SS_094. Tree of 5 sequenced nasal colonies. Colonisation was by ST5 strain. All colonies are genetically identical at core genome level. Tree was rooted using N315 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

No unique SNPs, insertions or deletions were identified in SS_094.

Control study participant SS_099

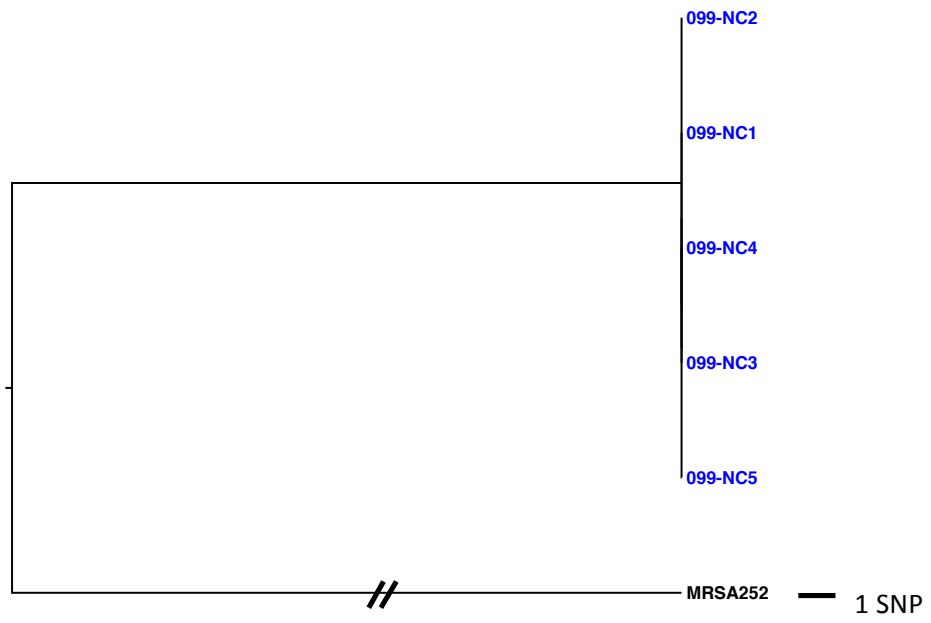

Figure S-8 Maximum likelihood core SNPs tree for control study ID SS_099. Tree of 5 sequenced nasal colonies. Colonisation was by ST30 strain. All colonies are genetically identical at core genome level. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

No unique SNPs, insertions or deletions were identified in SS_094.

Control study participant SS_141

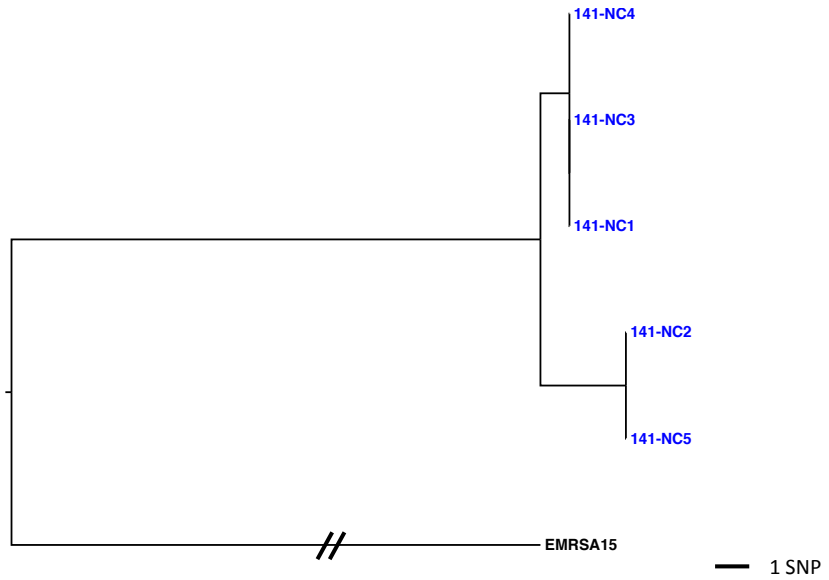

Figure S-9 Maximum likelihood core SNPs tree for control study ID SS_141. Tree of 5 sequenced nasal colonies. Colonisation was by ST22 strain. Tree was rooted using EMRSA15 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-5-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_141. Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in EMRSA15 reference (HO 5096 0412; accession number: HE681097). NS– non-synonymous.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
265069	A->G	NS	SAEMRSA1502110 putative zinc-binding dehydrogenase	Y->C	NC1/3/4
1042957	A->C	NS	pdhC SAEMRSA1509250 (dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex)	Y->S	NC1/3/4
2418873	C->A	NS	SAEMRSA1522430 (putative permease)	L->I	NC1/3/4

Table S-5-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_141. Unique indels in the core genome differentiating sequenced colonies. Base position in relation to reference assembly used for mapping (141-NC4) and region/ gene are relative to position in EMRSA15 reference (HO 5096 0412; accession number: HE681097). (+): Base insertion also indicated by upper case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
1589070	NC1/2/4	(+)C	present in *FnbpA* SAEMRSA1523990	frameshift

Control study participant SS_147

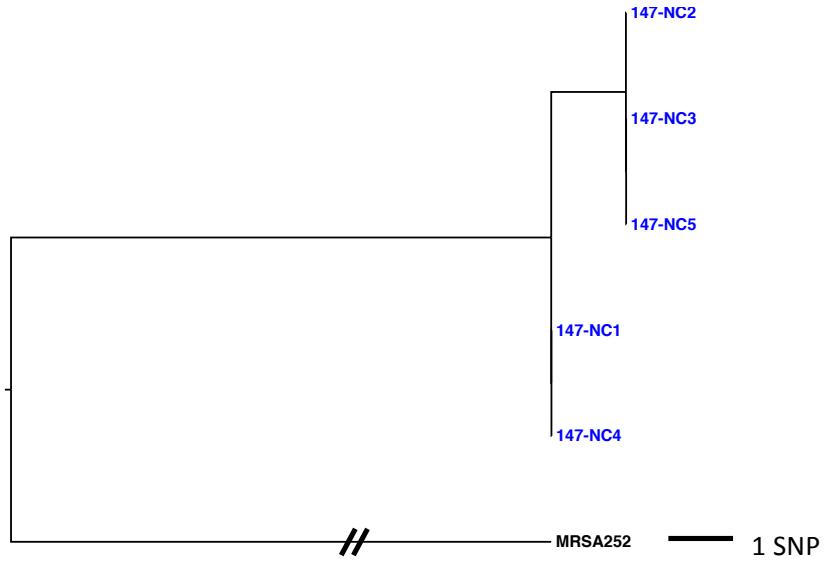

Figure S-10 Maximum likelihood core SNPs tree for control study ID SS_147. Tree of 5 sequenced nasal colonies. Colonisation was by ST30 strain. All colonies are genetically identical at core genome level. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-6 Within host-heterogeneity identified in nasal carriage control Study ID SS_147. Single nucleotide polymorphisms in core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in MRSA252 reference (accession number: BX571856). S- synonymous.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
506374	A->G	S	SAR0473 sugar-specific PTS transport system, IIBC component		NC2/3/5

No unique insertions or deletions were identified in SS_147.

Control study participant SS_153

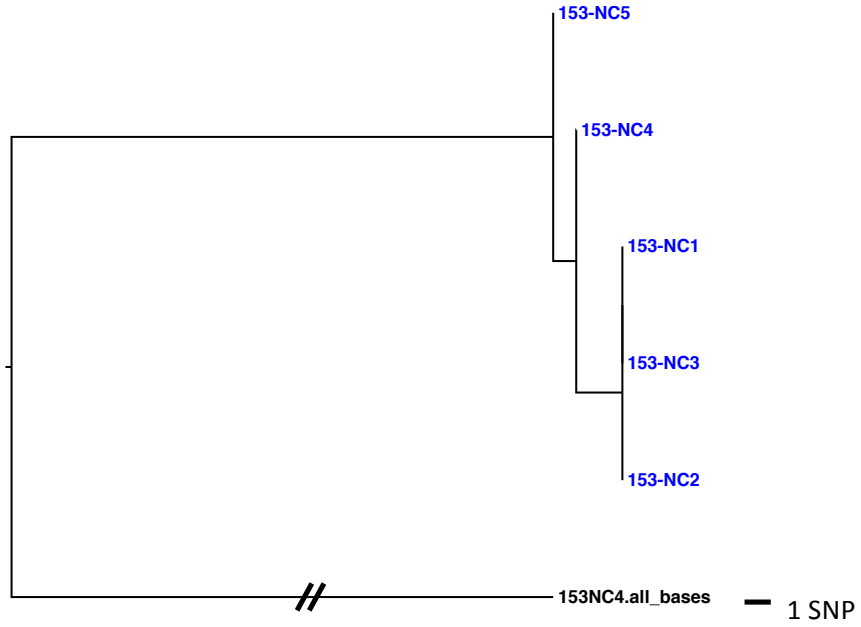

Figure S-11 Maximum likelihood core SNPs tree for control study ID SS_153. Tree of 5 sequenced nasal colonies. Colonisation was by ST15 strain. All colonies are genetically identical at core genome level. Tree was rooted using self-assembly of nose colony 4 (153NC4). SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-7 Within host-heterogeneity identified in nasal carriage control Study ID SS_153. Single nucleotide polymorphisms in core genome differentiating sequenced colonies. Base position and region are in relation to self-colony used for mapping (153NC4) with annotation with annotation transferred from MSSA476. NS– non-synonymous; S- synonymous.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
116038	T->C	NS	SAS1161 ATP-dependent DNA helicase	V->A	NC5
1437502	C->T	NS	SAS1346 putative membrane protein	V->I	NC1/NC3/ NC2
2507851	G->A	S	SAS0684 putative sulfatase		NC1/ NC3/ NC2

No unique insertions or deletions were identified in SS_153.

Control study participant SS_157

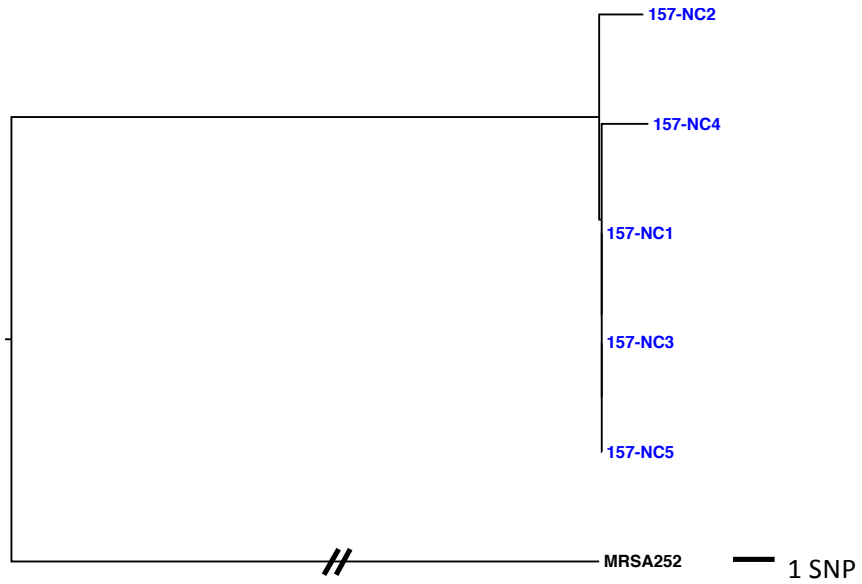

Figure S-12 Maximum likelihood core SNPs tree for control study ID SS_157. Tree of 5 sequenced nasal colonies. Colonisation was by an ST30 strain. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

Table S-8 Within host-heterogeneity identified in nasal carriage control Study ID SS_157. Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in MRSA252 reference (accession number: BX571856). NS– non-synonymous; S- synonymous.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
434094	C->T	NS	SAR0398- alkyl hydroperoxide reductase subunit F	A->T	NC4
2240292	T->C	S	SAR2168 putative helicase		NC2

No unique insertions or deletions were identified in SS_157.

Control study participant SS_250

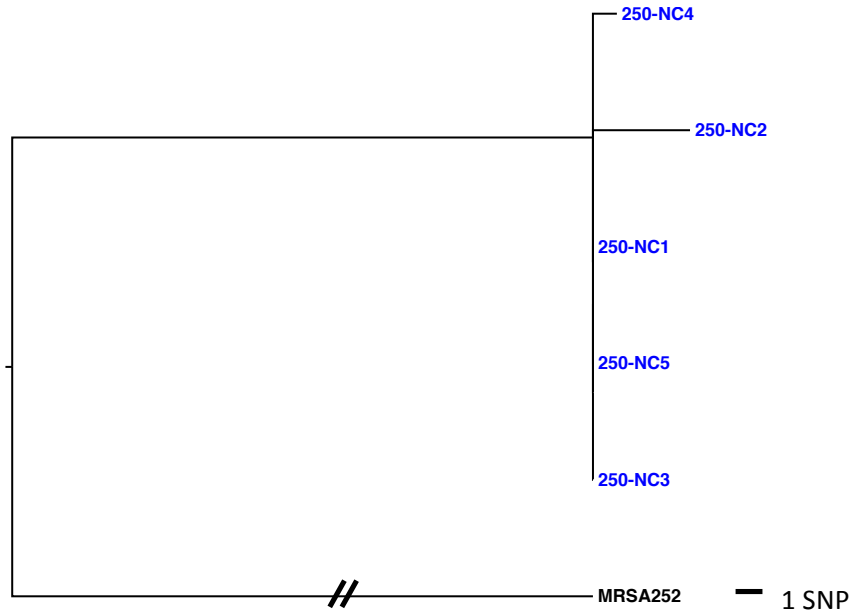

Figure S-13 Maximum likelihood core SNPs tree for control study ID SS_250. Tree of 5 sequenced nasal colonies. Colonisation was by an ST30 strain. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC)

Table S-9-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_250. Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in MRSA252 reference (accession number: BX571856). NS– non-synonymous; STOP- premature stop codon.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
1008575	C->T	NS	SAR0964 *trpS* putative tryptophanyl-tRNA synthetase	A->T	NC4
307160	G->T	NS	SAR0263 putative PTS transport system protein	D->Y	NC2
343726	C->T	NS	SAR0302 putative formate/nitrite transporter	G->S	NC2
1657450	G->A	NS	SAR1585 *malR* maltose operon transcriptional repressor	L->F	NC2
2344765	G->A	STOP	SAR2263 (putative membrane protein)	Q->*	NC1/3/4/5

Table S-9-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_250. Unique indels in the core genome differentiating sequenced colonies. Base position in relation to reference assembly used for mapping (250-NC4) and region/ gene are relative to position in MRSA252 reference genome (accession number: BX571856). (+): Base insertion also indicated by upper case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
483431	NC1/2/5/3	(+) T	Between SAR0931 (putative membrane protein) and SAR0932 (putative transposase)	Intergenic; no predicted consequence

Control study participant SS_268

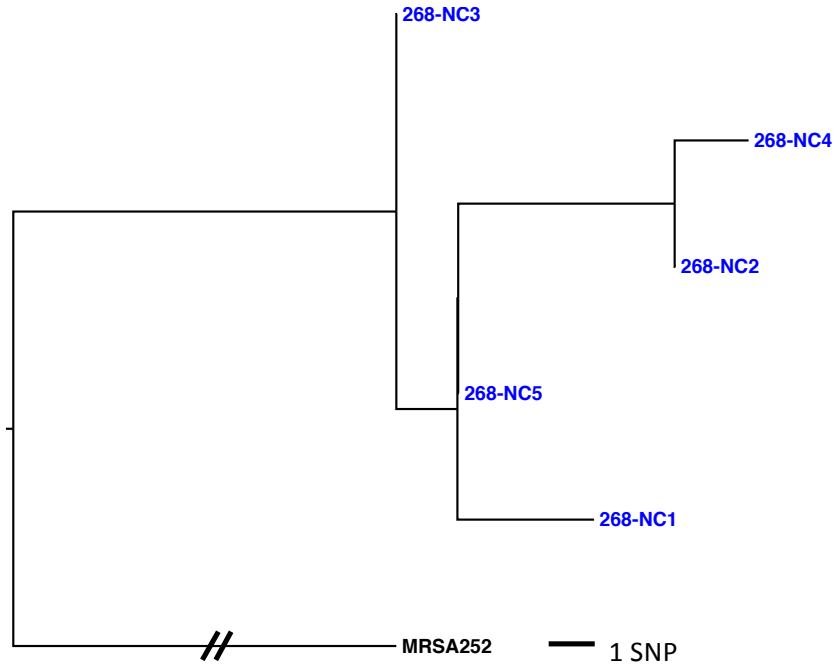

Figure S-14 Maximum likelihood core SNPs tree for control study ID SS_268. Tree of 5 sequenced nasal colonies. Colonisation was by an ST2889 strain. Tree was rooted using MRSA252 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC)

Table S-10-1 Within host-heterogeneity identified in nasal carriage control Study ID SS_268. Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are relation to position in MRSA252 reference (accession number: BX571856). NS– non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
1726183	A->G	NS	SAR1658 *grpE* GrpE protein Hsp-70 cofactor	M->T	NC3
1136521	C->T	NS	SAR1088 putative pyruvate carboxylase	P->S	NC2/4
1478391	G->A	NS	SAR1420 conserved hypothetical protein	T->I	NC4
2625727	T->A	NS	SAR2544 ABC transporter ATP-binding protein	I->F	NC1
2839685	T->G	S	SAR2734 *sasA S.aureus* surface protein A		NC1/5
1311725	C->T	S	SAR1251 (conserved hypothetical protein)		NC2/4
2792697	C->T	I	SAR2700 (putative membrane protein) and SAR2699 (putative glutathione peroxidase)		NC2/4
1745033	A->G	I	Within hypothetical phage protein (SAR1682)		NC1

Table S-10-2 Within host-heterogeneity identified in nasal carriage control Study ID SS_268. Unique indels in the core genome differentiating sequenced colonies. Base position in relation to reference assembly used for mapping (250-NC4) and region/ gene are relative to position in MRSA252 reference genome (accession number: BX571856). (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
1663853	NC1	(-) a	In SAR0264 bglA 6-phospho-beta-glucosidase	Frameshift with truncation

Control study participant SS_303

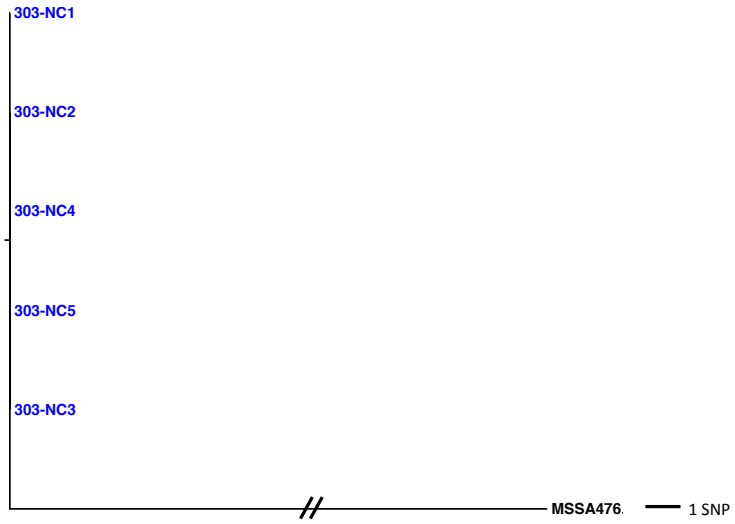

Figure S-15 Maximum likelihood core SNPs tree for control study ID SS_303. Tree of 5 sequenced nasal colonies. Colonisation was by an ST1 strain. Tree was rooted using MSSA476 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough). Branch labels are numbered according to the sequenced nasal colony (NC).

No unique SNPs, insertions or deletions were identified in SS_303.

Appendix B: Chapter 5 Supplementary data

Staphylococcus aureus associated with Atopic Eczema prospective case study

The following results are the characterisation of the remaining 5 AE cases, which were recruited as part of the study presented in Chapter 5. For each individuals sequenced colonies is a representative phylogeny, and where identified a table detailing the SNPs and indels arising within the AE cases samples.

Patient 2 (study ID PSAE002)

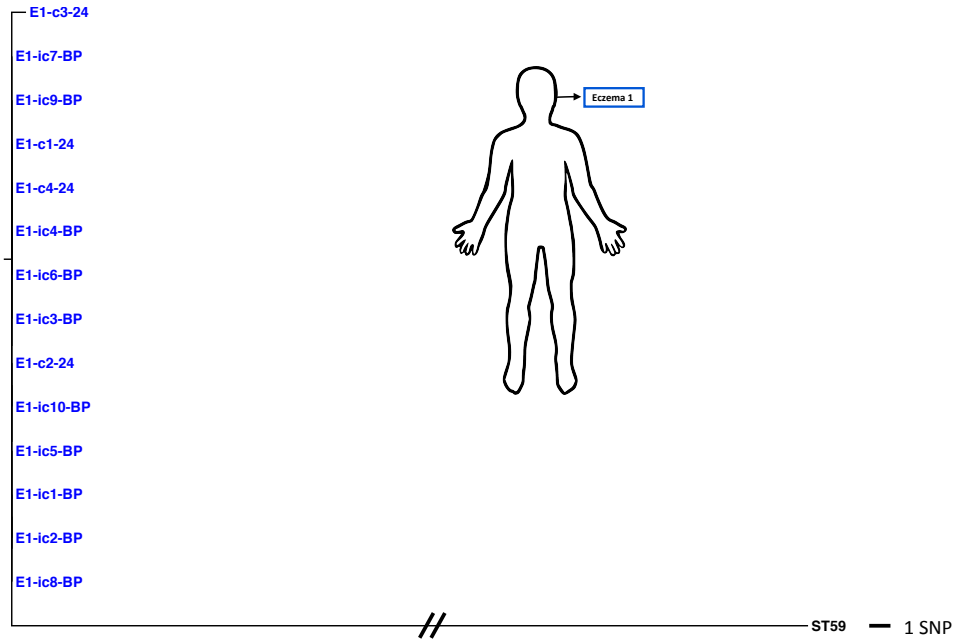

Figure S-16 Maximum likelihood core SNPs tree for Patient 2 (PSAE002). Tree of 14 sequenced colonies from single eczema affected site. Colonisation was by an ST59 strain. Colonies were isolated on differing selection agar labelled (**24**- staph brilliance; **BP**- Baird Parker). Body diagram illustrates sampling site, branch label colouring corresponds to sampling site. Tree was rooted using ST59 reference. Branch labels are numbered according to the sequenced and media used. All were derived without enrichment (labelled **ic**- initial colony). SNP bar is indicated for scale (not applicable to root branch with strikethrough).

Table S-11 Within-host heterogeneity identified in Patient 2 (PSAE002). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position in ST59 reference genome (MO13; accession number: CP003166). NS– non-synonymous.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
1620196	G->T	NS	M013TW_1550 (Glycine dehydrogenase (glycine cleavage system P2 protein)	T->N	E1-c3-24

No unique insertions or deletions were identified in Patient 2.

Patient 3 (study ID PSAE003)

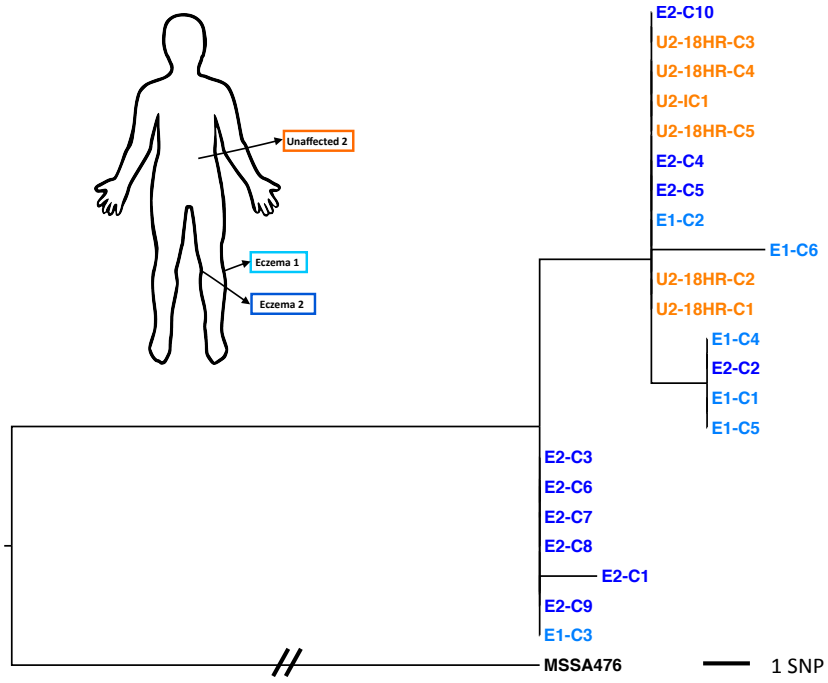

Figure S-17 Maximum likelihood core SNPs tree for Patient 3 (PSAE003). Tree demonstrating phylogenetic relationship of 22 sequenced colonies from sub-sampling of a single eczema affected site and one unaffected skin site. Body diagram illustrates sampling site, branch label colouring corresponds to sampling site. Colonisation was by an ST1 strain. Tree was rooted using MSSA476 reference. Branch labels: E- eczema site, U- unaffected site, C- colony number from body site. Colonies derived without enrichment are labeled IC - initial colonies; obtained by enrichment- 18 HR. SNP bar is indicated for scale (not applicable to root branch with strikethrough).

Table S-12-1 Within-host heterogeneity identified in Patient 3 (PSAE003). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position in ST1 reference genome (MSSA476; accession number: BX571857). NS– non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
200050	C->T	NS	SAS0170 *hsdR* putative type I restriction enzyme	T->M	E2-Col10, U2-18HRCol3, U2-18HRCol4, U2-IC1, U2-18HRCol5, E2-Col4, E2-Col5, E1-Col2, E1-Col6, U2-18HRCol2, U2-18HRCol1, E1-Col4, E2-Col2, E1-Col1, E1-Col5
1233332	G->T	NS	SAS1155 *rsgA* conserved hypothetical protein	E->D	E2-Col1
1604228	T->G	NS	SAS1480 *ComGC* competence protein	D->A	E1-Col6
1681182	C->T	S	SAS1559 putative oxygenase		E1-Col6
1773531	G>A	S	SAS1640 *tpx putative* thiol peroxidase		E1Col4, E2Col2, E1Col1, E1Col5
694264	C->T	I	Between SAS0614 (ferrichrome transport permease) and SAS0615 (putative dihydroxyacetone kinase)		E2-Col10, U2-18HR-Col3, U2-18HR-Col4, U2-IC1, U2-18HR-Col5, E2-Col4, E2-Col5, E1-Col2, E1-Col6, U2-18HR-Col2, U2-18HR-Col1, E1-Col4, E2-Col2, E1-Col1, E1-Col5

Table S-12-2 Within host-heterogeneity identified in Patient 3 (PSAE003). Unique indels in the core genome differentiating sequenced colonies. Base position in relation to reference assembly used for mapping (PSAE003_E1_C3) and region/ gene are relative to position in MSSA476 reference genome (accession number: BX571856). (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
834677	E2C10, U218HRC3, U218HRC4, U2C1, U218HRC5, E2C4, E2C5, E1C2, E1C6, U218HRC2, U218HRC1, E1C4, E2C2, E1C1, E1C5	(-) t	SAS1331- transposase fragment	Nil

Patient 6 (study ID PSAE006)

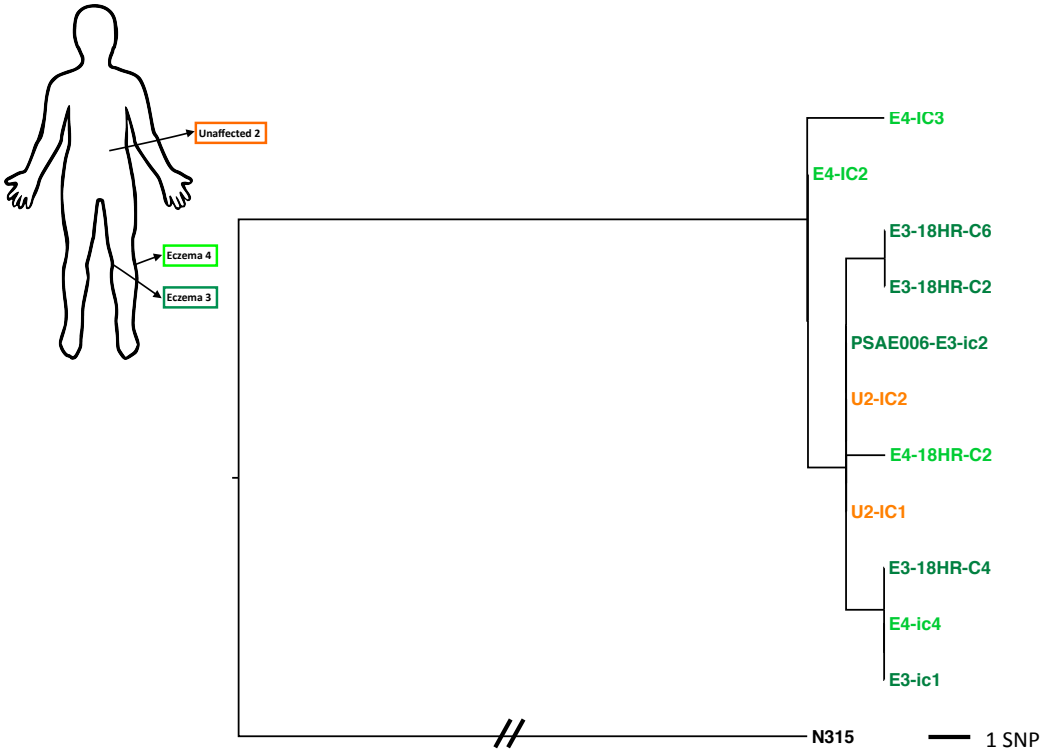

Figure S-18 Maximum likelihood core SNPs tree for Patient 6 (PSAE006). Tree demonstrating phylogenetic relationships of 11 sequenced colonies from sub-sampling of a single eczema affected site and one unaffected skin site. Colonies were obtained by both pre- (labeled ic- initial colony) and post-enrichment (labeled -18) culture methods. Body diagram shows sampling site, branch label coloring corresponds to body site colony was obtained from. Colonisation was by an ST5 strain. Branch labels: E- eczema site, U- unaffected site, C- colony number from body site. Colonies derived without enrichment are labeled IC - initial colonies; obtained by enrichment- 18 HR. Tree was rooted using N315 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough).

Table S-13 Within-host heterogeneity identified in Patient 6 (PSAE006). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position in ST5 reference genome (N315; accession number: BA000018). NS– non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
404620	G->A	NS	SA0345 conserved hypothetical protein	H->Y	E4-ic3
2547513	T->A	NS	SA2272 hypothetical protein	E->V	E4-ic4, E3-ic1, E3-18HR-col4
1295102	G->A	NS	SA1138 *mutL* DNA mismatch repair protein	E->K	E4-18HR-col2
1203913	C->T	S	SA1063 cation efflux superfamily serine threonine protein kinase		E4-ic4, E3-ic1, E3-18HR-col4, E3-18HR-col6, E3-18HR-col2, E3-ic2, U2-ic2, E4-18HR-col2, U2-ic1
499900	C->G	S	SA0433 (alpha-glucosidase)		E3-18HR-col6, E3-18HR-col2
497418	A->G	I	Between SA0431 (*gltD*) and SA0432 (*treP*)		E4-ic3

No unique insertions or deletions were identified in Patient 6.

Patient 9 (study ID PSAE009)

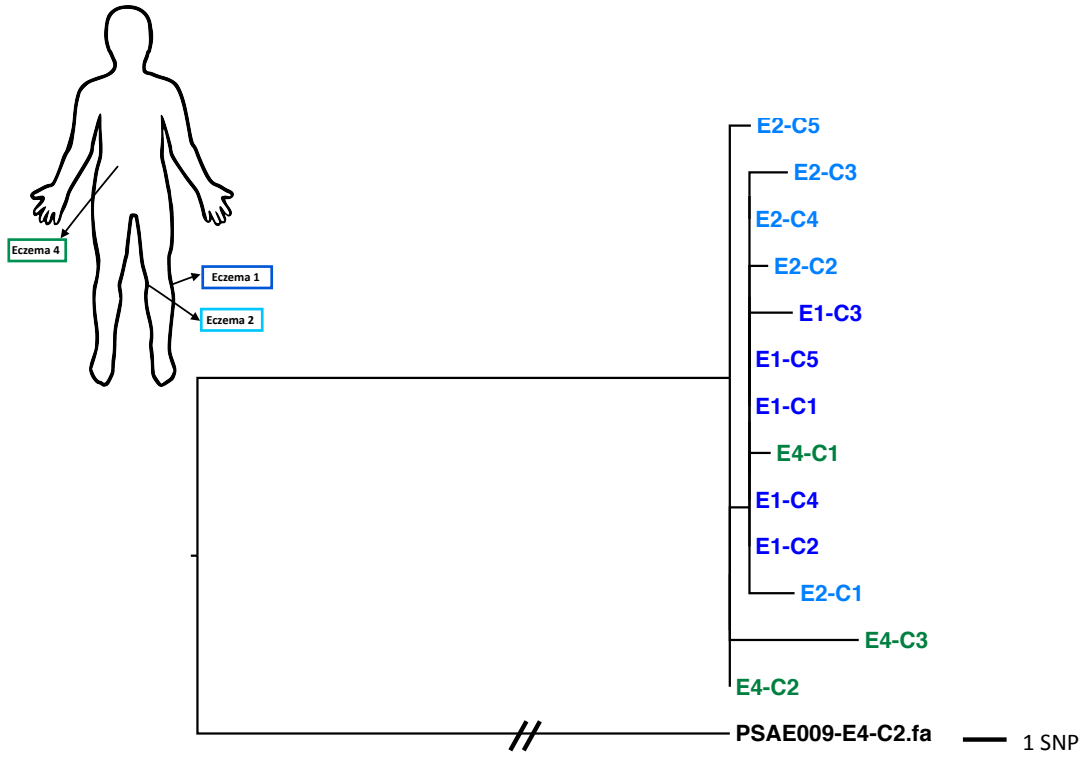

Figure S-19 Maximum likelihood core SNPs tree for Patient 9 (PSAE009). Tree demonstrating phylogenetic relationship of 13 sequenced colonies from two eczema-affected sites. Colonisation was by an ST2867 strain. Body diagram shows sampling site, branch label coloring corresponds to body site colony was obtained from. Branch labels: **E1-** medial border of single eczema site; **E2-** lateral border of single eczema site. Tree was rooted using self-assembly PSAE009-E4-C2. SNP bar is indicated for scale (not applicable to root branch with strikethrough).

Table S-14-1 Within-host heterogeneity identified in Patient 9 (PSAE009). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position in self-assembly reference with annotation transferred from MSSA476 (accession number: BX571856). NS– non-synonymous; S- synonymous; I- intergenic.

Base position	Base change	SNP type	Region/ gene	AA change	Colonies
420123	T->C	NS	SAS0201 *pflB* formate acetyltransferase	R->G	E4C3
1126393	C->G	NS	SAS1377 *ebH* very large surface anchored protein	A->G	E4C3
1809863	A->G	NS	SAS1342 *msa* putative membrane protein	D->G	E4C3
2732624	A->G	NS	SAS0833 putative membrane protein	Y->C	E4C3
2672196	T->C	NS	SAS0767 hypothetical protein	I->V	E2-C3, E2-C4, E2-C2, E1-C3, E1-C5, E1-C1, E4-C1, E1-C4, E1-C2, E2-C1
17069	G->A	NS	SAS2188 putative transcriptional regulator (AraC family)	S->N	E1C3
1934970	C->T	NS	SAS1431 *ResE sensor* kinase	E->K	E1C3
843559	A->T	NS	SAS2401 *fbp* Fructose-1-6-bispphophatase	P->R	E2-C3, E2-C4
102277	G->T	S	SAS2474 conserved hypothetical protein fructosamine kinase		E2-C3, E2-C4
833218	A->T	S	SAS2420 putative membrane protein		E2C2
644954	T->C	S	SAS2080 hypothetical protein		E2C5
951199	G->T	I	Between SAS1952 (ABC transporter ATP-binding protein and SAS1951 (conserved hypothetical protein)		E4C3
2455498	A->G	I	Between SAS1165 (3-oxoacyl-[acyl-carrier protein] reductase) and SAS1166 (acyl carrier protein)		E4C3

Table S-14-2 Within-host heterogeneity identified in Patient 9 (PSAE009). Unique indels in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position in self-assembly reference with annotation transferred from MSSA476 (accession number: BX571856). (-): base deletion also indicated by lower case lettering.

Base position	Colonies	Base change	Region/ gene	Predicted consequence
658556	E1C3	(-) gcgaaaacggt	SAS2067- ATP binding protein involved in chromosome partitioning	11 bp deletion; truncated protein
1618083	E4 C1	(-) a	SAS2057 PTS system, mannitol specific IIBC component	Truncated protein

Patient 10 (study ID PSAE010)

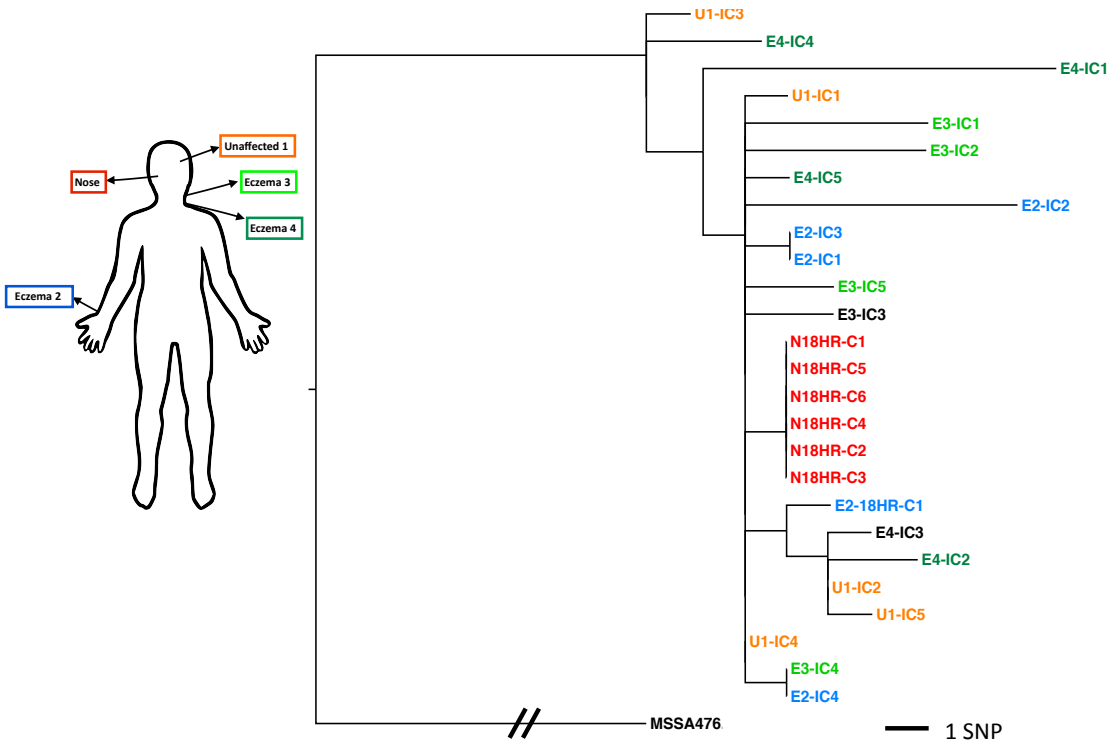

Figure S-20 Maximum likelihood core SNPs tree for Patient 10 (PSAE010). Phylogenetic tree of 26 sequenced colonies 4 body sites. Colonisation was by an ST1 strain. Body diagram shows sampling site, branch label coloring corresponds to body site colony was obtained from. Branch labels: E- eczema site, U- unaffected site, N- nose, C- colony number from body site. Colonies derived without enrichment are labeled IC - initial colonies; obtained by enrichment-18 HR. Tree was rooted using MSSA476 reference. SNP bar is indicated for scale (not applicable to root branch with strikethrough).

Table S-15-1 Within-host heterogeneity identified in Patient 10 (PSAE010). Single nucleotide polymorphisms in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position MSSA476 reference (accession number: BX571856). NS– non-synonymous; S- synonymous; I- intergenic.

Patient 10 SNPs					
Base position	**Base change**	**SNP type**	**Region/ gene**	**AA change**	**Colonies**
1741005	T->C	NS	SAS1618 DNA polymerase I	S->G	E2-18HR-C1
1239972	C->A	NS	SAS1161 *RecG* ATP-dependent DNA helicase	A->E	E4-IC2
564735	A->G	NS	SAS0494 *NusG* transcription antitermination protein	Q->R	N18HR-C1, N18HR-C5, N18HR-C6, N18HR-C4, N18HR-C2, N18HR-C3
1829799	T->C	NS	SAS1682 *sasC* putative surface anchored protein	I->M	E3-IC1
192515	C->T	NS	SAS0164 *ptsG* glucose-specific PTS transporter protein, IIABC component	G->D	E3-IC2
2469310	A->T	NS	SAS2309 *sbi* IgG-binding protein	E->V	E3-IC2
2469316	T->A	NS	SAS2309 *sbi* IgG-binding protein	V->D	E3-IC2
2778719	T->G	NS	SAS2578 *cna collagen* adhesin precursor	K->T	E4-IC5
1437876	C->T	NS	SAS1341 diaminopimelate decarboxylase	S->L	E2-IC2
1468016	T->C	NS	SAS1374 conserved hypothetical protein	H->R	E2-IC3, E2-IC1
208273	A->G	NS	SAS0177 putative membrane protein	H->R	E3-IC5
389724	G->A	NS	SAS0339 putative GTP-binding protein	V->I	E3-IC3
1863974	G->A	NS	SAS1716 O-succinylbenzoate synthase	L->F	E3-IC3
971178	G->A	NS	SAS0882 sodium: alanine symporter family protein	G->D	U1-IC1
80763	A->G	NS	SAS0070 *plc* 1-phosphatidylinositol phosphodiesterase precursor	I->V	E4-IC1
620557	T->C	NS	SAS0538 phosphomethylpyrimidine kinase	D->G	E4-IC1
867412	T->G	NS	SAS0785 *sufD* FeS cluster assembly protein	F->L	E4-IC1
914840	C->T	NS	SAS0833 hypothetical membrane protein	P->S	E4-IC1

1805141	G->C	NS	SAS1664 putative membrane protein	A->G	E4-IC1
2252001	G->A	NS	SAS2081 putative membrane protein	T->I	E4-IC1
322372	A->G	NS	SAS0276 conserved hypothetical protein	H->R	U1-IC3
1472333	A->T	NS	SAS1377 *ebh* very large surface protein	I->I	E3-IC1
1283703	T->A	S	SAS1198 *polC* DNA polymerase III PolC-type		E2-18HR-C1, E4-IC3, E4-IC2, U1-IC2, U1-IC5
2330729	T->C	S	SAS2175 acyl-coA dehydrogenase		E4-IC3, E4-IC2, U1-IC2, U1-IC5
2712271	G->A	S	SAS2525 putative exported protein		E4-IC2
152495	A->G	S	SAS0131 *cap8H* capsular polysaccharide synthesis enzyme		E2-IC2
1187325	C->T	S	SAS1115 penicillin-binding protein 1		E2-IC2
1963820	A->G	S	SAS1801 putative aminopeptidase		E2-IC2
2275858	A->G	S	SAS2103 hyaluronate lyase precursor 2		E2-IC2
1813839	G->A	S	SAS1671 putative aminopeptidase		U1 IC3, E4IC1
1301163	A->G	S	SAS1210 putative DNA translocase FtsK/SpoIIIE family protein		E3-IC5
2634395	C->T	S	SAS2453 *oatA* O-acetylytransferase OatA		E3-IC4, E2-IC4, U1-IC4, E2-18HR-C1, E4-IC3, E4-IC2, U1-IC2, U1-IC5, N18HR-C1, N18HR-C5, N18HR-C6, N18HR-C4, N18HR-C2, N18HR-C3, E3-IC1, E3-IC2, E4-IC5, E2-IC2, E2-IC3, E2-IC1, E3-IC5, E3-IC3, U1-IC1

1773369	A->G	S	SAS1640 *tpx* putative thiol peroxidase	E4-IC1
907009	A->G	I	Between SAS0827 (ornithine aminotransferase) and SAS0828 (putative NAD-specific glutamate dehydrogenase)	E3-IC4, E2-IC4
1344721	C->T	I	Between SAS1254 (hypothetical protein) and SAS1255 (hypothetical protein)	E4-IC3
721464	A->T	I	Between SAS0641 (putative acetyltransferase) and SAS0642 (putative lipoprotein)	U1-IC5
2756228	C->G	I	Between SAS2555 (intercellular adhesion protein C) and SAS2556 (lipase precursor)	E3-IC1
2756249	A->T	I	Between SAS2555 (intercellular adhesion protein C) and SAS2556 (lipase precursor)	E3-IC1
678744	G->A	I	Between SAS0601 (putative membrane protein) and SAS0602 (teichoic acid biosynthesis protein)	E3-IC2
904286	A->G	I	Between SAS0825 (putative S1 RNA binding domain) and SAS0826 (NADH oxidase family protein)	E2-IC2
1464511	T->C	I	Between SAS1369 (dihydrofolate reductase type I) and SAS1370 (thymidylate synthase)	E4-IC1
797591	C->T	I	Between SAS0714 (conserved hypothetical protein) and SAS0715 (putative helicase)	E4-IC4
1542831	T->C	I	Between SAS1417 (putative 30S ribosomal protein S1) and SAS1418 (cytidylate kinase)	E4-IC4
2758687	G->A	I	Between SAS2556 (lipase precursor) and SAS2557 (putative histidine biosynthesis bifunctional protein)	E4-IC4

Table S-15-2 Within-host heterogeneity identified in Patient 10 (PSAE010). Unique indels in the core genome differentiating sequenced colonies. Base position and region/ gene are in relation to position MSSA476 reference (accession number: BX571856). (+): Base insertion also indicated by upper case lettering; (-): base deletion also indicated by lower case lettering.

Base position	Base change	Region/ gene	Predicted consequence	Colonies

1032266	(-) A	Intergenic region between SAS0428 (LysR family regulatory protein) and SAS0429 (glutamate synthase large subunit)	Potentially in translation start site; in putative RBS - not obvious effect on translation expression	E3IC1
1243646	(-) gtata	Intergenic between SAS2404 (putative dioxygenase) and SAS2405 (MarR family regulatory protein)	Upstream of translation start; RBS disrupted; potential detrimental effect on expression of the gene	E3IC2
1126776	(-) t	Present in SAS2515 putative esterase	Frameshift; F->H; truncated variant; over half of protein un-translated	U1IC2
654616	(+) T	Intergenic region between SAS1127 (isoleucyl-tRNA synthetase) and SAS1128 (glyoxalase/bleomycin resistance protein)	Intergenic no predicted effect	E218HRC1, E2IC2, E2IC4, E3IC1, E3IC2, E3IC4, E3IC5, E4IC2, E4IC3, E4IC1, E4IC5, U1CI, U1C2, U1C3, U1C4, U1C5
2368412	(-) a	SAS2322 hypothetical protein	No consequence; doubtful CDS	All except nose colonies
650742	(-) a	SAS1125 S4 RNA binding protein	Truncation; 48 AA's reduction	E3 IC4, E2IC4, E3IC5, E4IC5